

Aus dem
Institut für Medizinische Statistik und Dokumentation
der Universität Erlangen-Nürnberg
Direktor: Professor Dr. L. Horbach

Analyse der empirischen Verteilungen von Präzisionsstandards im klinisch-chemischen Labor

Inaugural - Dissertation
zur Erlangung der Doktorwürde
der Medizinischen Fakultät
der Friedrich-Alexander-Universität
Erlangen - Nürnberg

vorgelegt von
Erwin Münch
aus Schweinfurt

1981 0463

Aus dem
Institut für Medizinische Statistik und Dokumentation
der Universität Erlangen - Nürnberg
Direktor: Professor Dr. L. Horbach

Analyse der empirischen Verteilungen von Präzisionsstandards im klinisch-chemischen Labor

Inaugural - Dissertation
zur Erlangung der Doktorwürde
der Medizinischen Fakultät
der Friedrich-Alexander-Universität
Erlangen - Nürnberg

vorgelegt von
Erwin Münch
aus Schweinfurt

Gedruckt mit Erlaubnis der Medizinischen Fakultät
der Universität Erlangen-Nürnberg



Dekan: Prof. Dr. V. Becker
Referent: Prof. Dr. L. Horbach

Tag der mündlichen Prüfung: 30.11.1981

Einleitung

1. Zielsetzung

Das Ziel dieses Projekts ist es, die

Ergebnisse der Untersuchung

zu präsentieren und

zu diskutieren.

Die Ergebnisse werden in

den folgenden Kapiteln

präsentiert und diskutiert.

Die Ergebnisse werden in

2. Ergebnisse

Die Ergebnisse der Untersuchung

sind in den folgenden

Tabellen dargestellt.

Die Ergebnisse der Untersuchung

sind in den folgenden

Texten dargestellt.

Die Ergebnisse der Untersuchung

sind in den folgenden

Texten dargestellt.

Die Ergebnisse der Untersuchung

sind in den folgenden

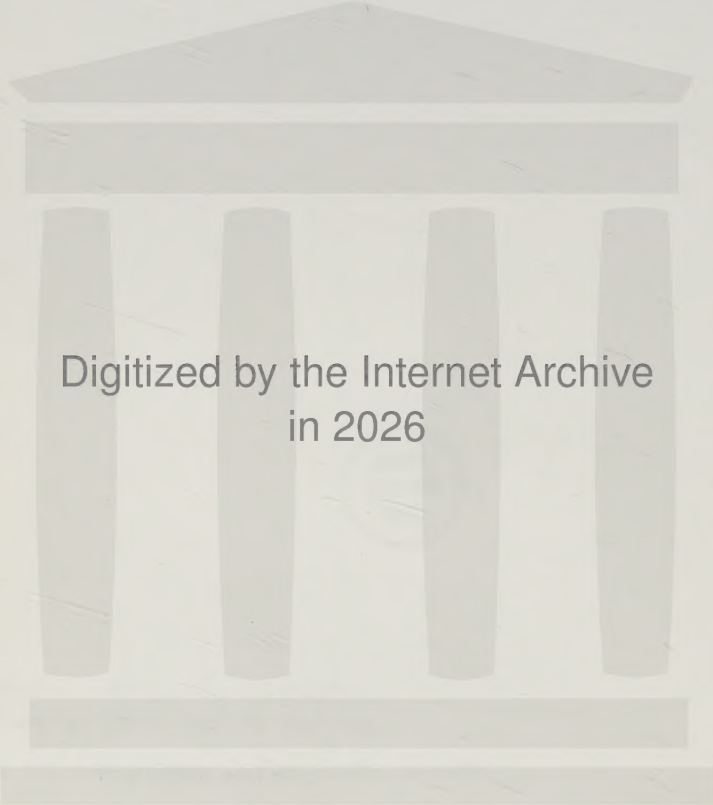
Texten dargestellt.

Die Ergebnisse der Untersuchung

sind in den folgenden

Texten dargestellt.

Meinen Großeltern



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41546216_0087

Inhaltsverzeichnis

	Seite
0. Einleitung	1
1. Grundlagen der Präzisionskontrolle im klinisch-chemischen Laboratorium	4
1.1 Der zufällige Fehler	4
1.2 Der systematische Fehler	10
1.3 Auswirkungen variierender Meßfehlervarianzen auf die Klassifizierung von Patientenmeßwerten	14
1.4 Meßfehlervariabilität und "biologische Variabilität"	18
1.5 Prinzip und Durchführung der Präzisionskontrolle	23
2. Fragestellung	26
3. Beschreibung des Datenmaterials	28
3.1 Gewinnung der Daten	28
3.2 Empirische Verteilungen	31
4. Test der empirischen Präzisionsstandardverteilungen auf Normalverteilung	41
4.1 Zur Auswahl eines Tests auf Normalverteilung	41
4.1.1 Ein nicht näher bezeichneter Test nach VAN DER WAERDEN	42
4.1.2 Der χ^2 -Anpassungstest	44
4.1.2.1 "Gängige Methode" (VAN DER WAERDEN)	46
4.1.2.2 Intervallgrenze als Mittelpunkt zwischen benachbarten Ausprägungen	47
4.2 Procedere bei der Durchführung des χ^2 -Anpassungstests mit Hinweisen auf das ausführende Fortran-Programm	47

	Seite
4.3 Einführung einer weitgehend verteilungsunabhängigen Toleranzgrenze für die Präzision	50
5. Diskussion	64
5.1 Ergebnisse und deren Interpretation	64
5.2 Verfahren bei Ablehnung der Hypothese der Normalverteilung der Präzisionsstandards	67
5.3 Beurteilung der Testergebnisse der empirischen Präzisionsstandardverteilungen auf Normalverteilung unter Berücksichtigung der Präzision	73
6. Zusammenfassung	77
7. Literaturverzeichnis	78

0. Einleitung

Die im klinisch-chemischen Labor ermittelten Konzentrationen von Elektrolyten und Metaboliten des menschlichen Organismus sowie die Aktivitäten von Enzymen sollen dem Arzt ein Bild von der Stoffwechselsituation des jeweiligen Patienten vermitteln. Quantitative chemische Analysen sind jedoch, wie alle Messungen, nicht frei von Fehlern. Insbesondere im klinisch-chemischen Labor, wo die Proben im allgemeinen wässrige Lösungen mehrerer Substanzen sind, wird die Ermittlung möglichst exakter Konzentrationsmeßwerte eines Elektrolyten oder Metaboliten bzw. die Bestimmung der Aktivität eines Enzymes dadurch erschwert, daß die Konzentration der zu messenden Substanz niedrig ist, während die Konzentration von Substanzen, die den Meßvorgang stören, hoch ist. Hinzu kommen die besondere Empfindlichkeit des biologischen Materials und die Vielzahl von Einflußfaktoren während der verschiedenen Stufen der Verarbeitung (z. B. Blutentnahme, Transport, Zusatz von Reagenzien, Zentrifugation et cetera). Die Gesamtheit dieser Verarbeitungskette und der dabei eingesetzten Apparaturen, von der Blutentnahme beim Patienten bis zur eigentlichen Messung der interessierenden Stoffwechselvariablen, sei Meßsystem genannt.

Um zu gewährleisten, daß die in diesem komplizierten Meßsystem ermittelten Werte von den tatsächlichen nicht allzu sehr abweichen, werden Richtigkeitskontrollen zur Erfassung von systematischen Abweichungen und Präzisionskontrollen zur Ermittlung von zufälligen Fehlern durchgeführt.

Aufgabe der vorliegenden Arbeit war es, einige wichtige Voraussetzungen der Präzisionskontrolle an einer Stelle im Meßsystem, nämlich der chemisch-quantitativen Analyse, aufgrund empirischer Daten zu untersuchen. Zur Verfügung standen insgesamt 100 Meßreihen aus dem klinisch-chemischen Zentrallabor der Universitäts-Klinik in Erlangen. Dabei handelt es sich um wiederholte Messungen sogenannter Präzisionskontrollproben, auch Präzisionsstandards genannt: Es sind dies Proben, die aus einem "Pool" gründlich gemischter Substanzen entnommen werden. Gleiche Substanzen verschiedener Proben weisen also die gleiche Konzentration auf.

Eingestreut in die Tagesserien der Patientenproben vermögen deren Meßresultate Aufschluß über die Präzision der Meßgeräte zu geben. Die hier untersuchten Meßreihen beziehen sich auf 31 verschiedene Stoffwechselvariable. Pro Variable liegen mehrere Meßreihen aus verschiedenen Untersuchungsmaterialien vor. Im allgemeinen basiert die Präzisionskontrolle auf der Annahme, daß die Meßwerte der Präzisionskontrollproben (Präzisionsstandards) normal verteilt sind. Die Bewertung von Abweichungen erfolgt unter Bezug auf das Modell der Normalverteilung, wobei eine Additivität der Fehlerkomponenten vorausgesetzt wird.

Die Präzisionskontrolle wird heute von den meisten klinisch-chemischen Labors durchgeführt. Systematische Überprüfungen, ob die Voraussetzung der Normalverteilung der Präzisionsstandardwerte erfüllt ist, sind nicht bekannt. In dieser Arbeit werden die empirischen Verteilungen der Präzisionsstandards einiger im Labor untersuchter Variabler auf Übereinstimmung mit Normalverteilungen geprüft. Weiterhin wird untersucht, ob der aus den Präzisionsstandards errechnete Meßfehler nicht zu

hoch ist (also ob die Präzision nicht zu gering ist). Ein zu diesem Zweck entwickeltes Computerprogramm sollte nicht nur dazu dienen, die Frage einer signifikanten Abweichung der empirisch ermittelten Verteilung von der Normalverteilung zu prüfen, sondern auch - durch Ausdrucken von Zwischenergebnissen der Berechnungen - Aufschluß über die Form der Abweichung zu geben.

1. Grundlagen der Präzisionskontrolle im klinisch-chemischen Laboratorium

In diesem Kapitel werden einige statistische Sachverhalte und Begriffe, die zur Darstellung der Ergebnisse unerlässlich sind, kurz erläutert.

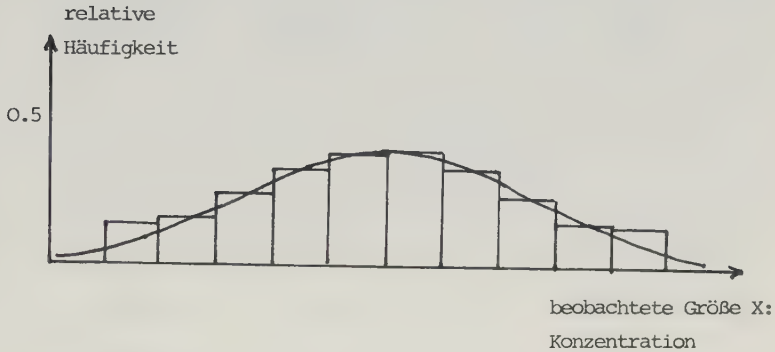
1.1 Der zufällige Fehler

Mißt man die Konzentration X einer in einer Probe befindlichen Substanz n -mal hintereinander und erhält so die Werte $x_1, \dots, \dots, x_n$, so sind diese im allgemeinen nicht identisch, sondern weisen eine durch zufällige Fehler bedingte Streuung auf. Die Gesamtheit der Werte $x_1, \dots, x_n$ nennt man auch Stichprobe aus der Grundgesamtheit aller möglichen Meßergebnisse dieser Konzentration unter den gleichen Meßbedingungen. Teilt man diese Werte in Klassen ein und trägt die Anzahl der Werte pro Klasse auf der Ordinate auf, so erhält man ein Histogramm, das einer empirischen Dichtefunktion $f_E(x)$ entspricht. Die kumulativen relativen Häufigkeiten ergeben die empirische Verteilung $F_E(x)$. Die beiden genannten Funktionen sind Treppenfunktionen. Durch Interpolation erhält man kontinuierliche Funktionen (siehe Abb. 1.1.1).

Die Verteilungen repräsentativer Stichproben von endlichem Umfang weichen in der Regel zufällig von der Verteilung der Grundgesamtheit ab, sie sind nur Schätzungen für die Verteilung der Grundgesamtheit. Diese Schätzungen sind umso genauer, je größer der jeweilige Stichprobenumfang ist.

Abb. 1.1.1

Graphische Darstellung einer empirischen Dichtefunktion (Histogramm) mit einer Interpolationskurve.



Graphische Darstellung der kumulativen relativen Häufigkeiten (empirische Verteilung) mit einer Interpolationskurve.

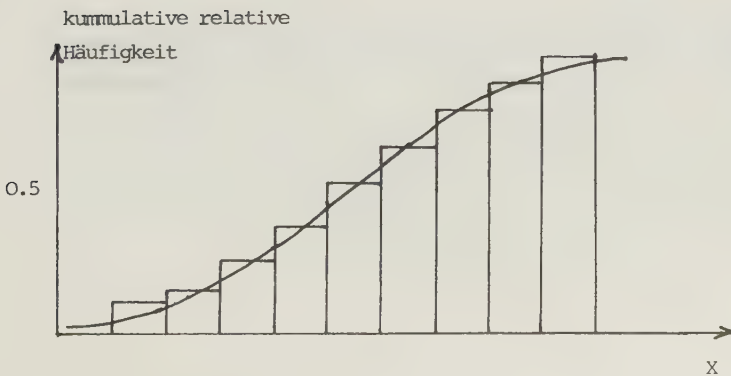
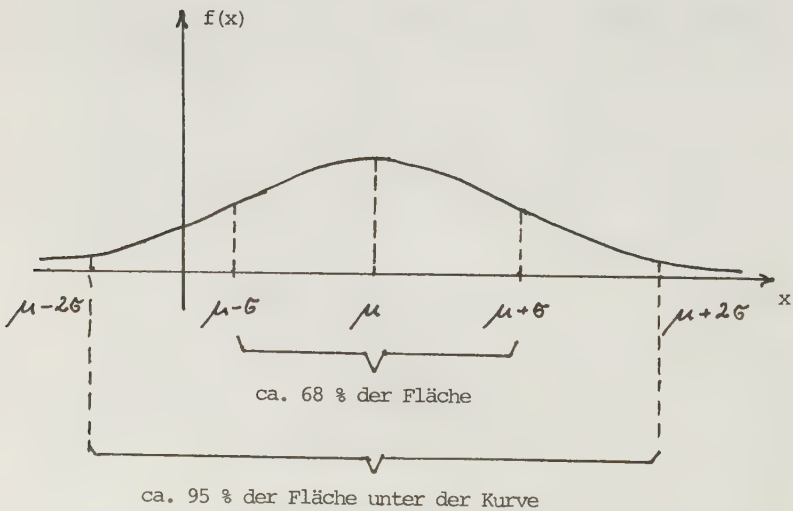
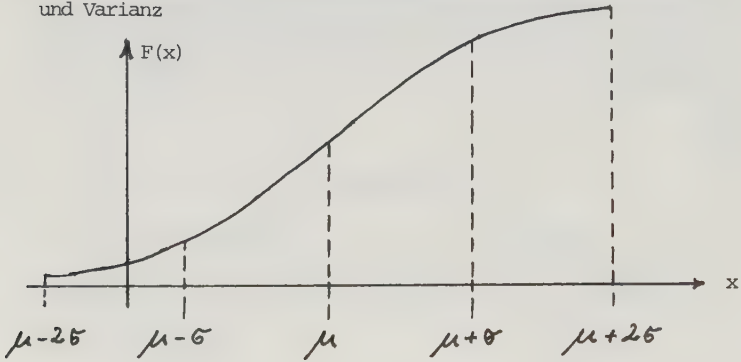


Abb. 1.1.2

Graphische Darstellung der Verteilungsfunktion $F(x)$ und der Dichtefunktion $f(x)$ einer Normalverteilung mit Mittelwert und Varianz



Die Verteilung aller möglichen Meßergebnisse einer bestimmten konstanten Konzentration ist charakteristisch für den Meßfehler. Man nennt sie auch Meßfehlerverteilung. In dieser Arbeit werden Funktionen, die sich auf den Meßfehler beziehen, mit dem Subscript "M" indiziert, also z. B. F_M (Verteilungsfunktion des Meßfehlers) und f_M (Dichtefunktion des Meßfehlers).

Wenn die Abweichungen der Konzentrationsmeßwerte von der tatsächlich in der Probe vorliegenden Konzentration zufälliger Natur sind, dann ist $f_M(x)$ die Dichtefunktion der Normalverteilung mit Mittelwert μ und Varianz σ^2 . Die Normalverteilung ist durch die Parameter μ und σ eindeutig bestimmt. Für die Normalverteilung gilt:

$$\int_{\mu-\sigma}^{\mu+\sigma} f(x) dx = F(\mu+\sigma) - F(\mu-\sigma) = 0.6827$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} f(x) dx = F(\mu+2\sigma) - F(\mu-2\sigma) = 0.9549$$

Die aus der Stichprobe errechenbaren Größen Mittelwert $\bar{x}$ und Standardabweichung s sind Schätzwerte für μ und σ . Die Standardabweichung wird als Dispersionsmaß zur Kennzeichnung der Präzision einer Messung benutzt. Abb. 1.1.2 zeigt die Dichtefunktion f_{M1} einer Normalverteilung mit einer kleineren Varianz σ_{M1}^2 (entspricht "kleinerem" Meßfehler) und die Dichtefunktion

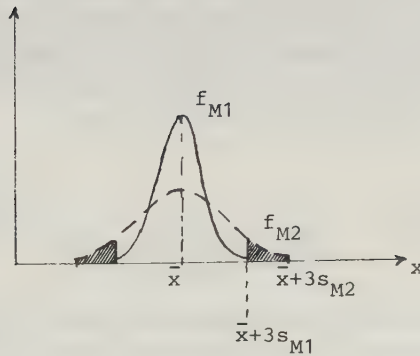
f_{M2} einer Normalverteilung mit einer größeren Varianz $\sigma_{M_2}^2$
(entspricht "größerem" Meßfehler).

Die empirische Meßfehlerverteilung, die aus den in einem Zeitintervall (t_1, t_2) gemessenen Werten ermittelt wird, stimmt im allgemeinen aber nur dann mit einer Normalverteilung überein, wenn in diesem Zeitintervall (t_1, t_2) der Meßfehler und damit auch μ und σ konstant bleiben. Dies wird in dieser Arbeit als Meßfehlerkonstanz bezeichnet. Eine Übereinstimmung zwischen Meßfehlerverteilung und Normalverteilung bei fehlender Meßfehlerkonstanz ist, allerdings mehr oder weniger theoretisch, nur dann denkbar, wenn die Änderung des Meßfehlers selbst eine normalverteilte Zufallsvariable ist, wenn diese vom Meßfehler stochastisch unabhängig ist und wenn deren "Periode" klein ist gegenüber dem Beobachtungszeitraum (t_1, t_2) .

Der Mittelwert als Dispersionsmaß spielt eine Rolle bei der Erkennung systematischer Abweichungen, der sogenannten Richtigkeitskontrolle (siehe 1.2).

Abb. 1.1.2

Dichtefunktion zweier Normalverteilungen mit unterschiedlicher Varianz.



f_{M1} = Dichte einer Normalverteilung mit Mittelwert $\bar{x}$ und Standardabweichung s_{M1}

f_{M2} = Dichte einer Normalverteilung mit Mittelwert $\bar{x}$ und Standardabweichung s_{M2} , $s_{M2} > s_{M1}$

1.2 Der systematische Fehler

Wie der Name schon sagt, verursacht ein systematischer Fehler - systematisch bei jeder Messung - die gleiche Abweichung des Meßresultates von der tatsächlich vorliegenden Meßgröße. Ursachen für einen systematischen Fehler im klinisch-chemischen Labor können zum Beispiel fehlerhafte Eichung eines Meßinstrumentes, falsche Konzentration eines Reagenz oder ähnliches sein. Abb. 1.2.1 zeigt die Verteilung der Meßergebnisse X einer tatsächlich vorliegenden Konzentration μ_0 , gemessen in einem hypothetischen Meßsystem, in dem es keinen zufälligen Fehler, sondern nur den systematischen Fehler $(\bar{x} - \mu_0)$ gibt. Die Wahrscheinlichkeitsfunktion dieser Verteilung ist nicht darstellbar. Sie entspricht dem "Dirac-Stoß"

$$\delta(x) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{x^2}{2\sigma^2}\right)}$$

In der Realität wird der systematische Fehler von zufälligen Fehlern überlagert. Werte wiederholter Messungen sind dann normalverteilt mit Mittelwert $\bar{x}$ und Standardabweichung s , wobei die mittlere Abweichung $d = \bar{x} - \mu_0$ der Meßergebnisse von der tatsächlich vorliegenden Größe μ_0 als Richtigkeit (Maß des systematischen Fehlers) und s als Präzision (Maß des zufälligen Fehlers) bezeichnet werden (siehe Abb. 1.2.2).

Allerdings führen nicht alle systematischen Fehler nur zu einer linearen Verschiebung der Verteilung der zufälligen Fehler auf der Abszisse, sondern, unter Umständen, auch zu Verzerrungen, und zwar dann, wenn der systematische Fehler von der zu messenden

Größe abhängt. Es ergeben sich in diesem Fall - unter der Voraussetzung, daß der zufällige Fehler dem systematischen "vorgeschaltet" ist - von einer Normalverteilung abweichende Verteilungen, wie die in Abb. 1.2.3 aufgeführte schiefe Verteilung zeigt. Solch eine Verteilung kann z. B. bei Einstellung einer falschen Inkubationstemperatur bei enzymkinetischen Methoden resultieren.

Im automatisierten Labor sind Kontrollen der Präzision und der Richtigkeit besonders wichtig. Im Gegensatz zu manuell durchgeführten Messungen sind Fehler nicht ad hoc erkennbar. Sie bleiben bestehen und wirken sich auf alle zukünftigen Messungen aus, wenn keine (Güte-)Kontrollen eingeschaltet werden.

Abb. 1.2.1

Graphische Darstellung einer Verteilung von Werten wiederholter Konzentrationsmessungen an einer Probe mit konstanter Konzentration μ_0 in einem hypothetischen Meßsystem, in dem es nur den systematischen Fehler $(x - \mu_0)$ gibt.

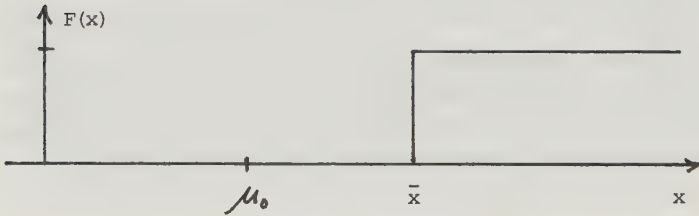


Abb. 1.2.2

Graphische Darstellung einer empirischen Dichtefunktion von Werten wiederholter Konzentrationsmessungen an einer Probe mit konstanter Konzentration μ_0 in einem Meßsystem, das mit dem systematischen Fehler $d = x - \mu_0$ und dem zufälligen Fehler (= Streuung) s behaftet ist.

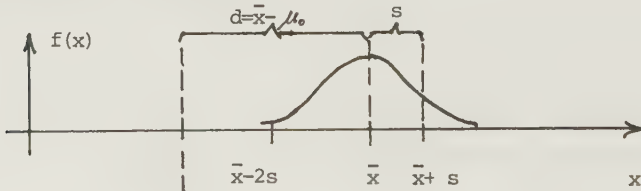
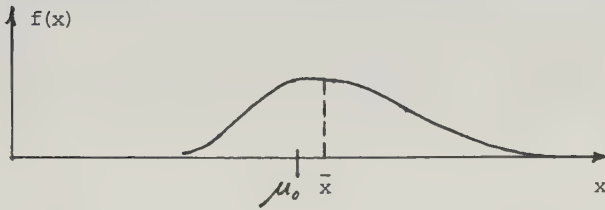


Abb. 1.2.3

Empirische Dichte von Meßwerten einer Kontrollprobe mit der Konzentration μ_0 in einem Meßsystem mit einem zufälligen und systematischen Fehler, wobei der zufällige Fehler "vor" dem systematischen auftritt und letzterer von der Konzentration X abhängt.



1.3 Auswirkungen variierender Meßfehlervarianzen auf die Klassifizierung von Patientenmeßwerten

Sinn der Präzisionskontrolle ist die Aufrechterhaltung eines gewissen Niveaus der Meßgenauigkeit und damit auch die Beschränkung der Fehlerrate bei der Zuordnung von Patienten in die beiden Klassen der "Normalen" und "Kranken" bezüglich einer Stoffwechselproduktkonzentration (Variablen) X. Ohne näher auf die Problematik der Definition der Begriffe "Normalität", "Gesundheit" und "Krankheit" einzugehen (siehe (20)), wird das Normalpersonenkollektiv mit dem Kollektiv der Gesunden sowie das "Nicht-Normalpersonenkollektiv" mit dem Kollektiv der Kranken gleichgesetzt.

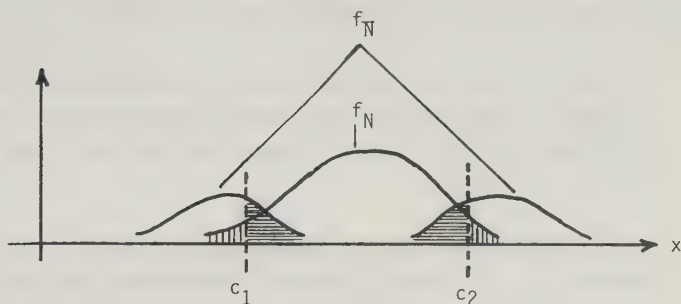
Die Klassifikation eines Patienten wird durch Vergleich des Meßwertes der Variablen X mit dem Normbereich dieser Variablen vorgenommen: Liegt der Meßwert des Patienten innerhalb des Normbereiches, dann wird dieser der Klasse der Normalen zugeordnet, andernfalls der der "Nicht-Normalen". Die Grenzen des Normbereiches einer Variablen X werden aus der Verteilung f_N der Meßwerte dieser Variablen eines Normalpersonenkollektives ermittelt. Im allgemeinen wird als untere Grenze c_1 des Normbereiches das 2,5-Perzentil dieser Verteilung und als obere Grenze c_2 das 97,5-Perzentil herangezogen. Aber auch die Beträge $\bar{x} - 3s$ bzw. $\bar{x} + 3s$ finden als untere bzw. obere Normbereichsgrenzen Verwendung.

Die obengenannte Zuordnungsvorschrift stellt also einen Klassifikator dar, der

$$\int_{-\infty}^{c_1} f_N(x) dx + \int_{c_2}^{+\infty} f_N(x) dx \quad \cdot 100 \% \text{ der}$$

Abb. 1.3.1

Graphische Darstellung der Fehlerrate bei der Zuordnung eines Patienten aus der Grundgesamtheit aller Patienten zur Klasse der "Normalen" beziehungsweise der "Nichtnormalen" an Hand des Meßwertes der Variablen X und deren Normbereich (c_1, c_2) .



f_N : Verteilung der Meßwerte der Variablen X im Normalpersonenkollektiv

$f_{\bar{N}}$: Verteilung der Meßwerte der Variablen X im Kollektiv der Nichtnormalen

Die schraffierte Fläche kennzeichnet den Anteil der falsch klassifizierten Patienten.

Patienten aus dem Normalpersonenkollektiv der Klasse der Nicht-Normalen zuordnet. Sei $f_{\bar{N}}(x)$ die Verteilung der Meßwerte der Variablen X im Kollektiv der Nicht-Normalen, sei $p(N)$ die a-priori-Wahrscheinlichkeit der Klasse der Normalen und sei $p(\bar{N})$ die a priori-Wahrscheinlichkeit der Klasse der "Nicht-Normalen", dann beträgt die Fehlerrate $\mathcal{G}$ dieses Klassifikators

$$[1] \quad \mathcal{G} = p(N) \cdot \left(1 - \int_{c_1}^{c_2} f_N(x) dx \right) + p(\bar{N}) \cdot \left(\int_{c_1}^{c_2} f_{\bar{N}}(x) dx \right)$$

(siehe Abb. 1.31). In die Verteilung $f_N(x)$ gehen sowohl die im Kollektiv der Normalen existierende Variabilität der Variablen X ein, als auch die Variabilität des Meßfehlers (analoges gilt für $f_{\bar{N}}(x)$).

Die Verteilung des Meßfehlers ist im allgemeinen eine Funktion $f_M(x, \mu_0)$, die vom Meßpunkt μ_0 (Konzentration der Probe, mit der die Meßfehlerverteilung ermittelt wird) abhängig ist (siehe Abb. 1.2.2). Im folgenden soll diese Abhängigkeit vernachlässigt werden, da sich das Prinzip der Auswirkung variierender Meßfehlervarianzen auf die Klassifikationsfehlerrate anhand einer Meßfehlerverteilung $f_M(x)$, die vom Meßpunkt μ_0 unabhängig ist, leichter darstellen läßt. Die Annahme der Unabhängigkeit der Meßfehlerverteilung vom Meßpunkt ist der Annahme der Konstanz von Präzision und Richtigkeit für alle μ_0 im Meßbereich äquivalent.

Sei $f_{NB}(x)$ die Verteilung der Variablen X im Kollektiv der Normalen und $f_M(y)$ die Verteilung des Meßfehlers Y und sei $f_M(y)$

unabhängig vom Meßpunkt μ_0 , dann ergibt sich die Verteilung der Meßwerte der Variablen X, die in dem mit dem Meßfehler Y behafteten Meßsystem gemessen wurden, als das Faltungsprodukt

$$[2] \quad f_N = f_{NB} * f_M = \int_{-\infty}^{+\infty} f_{NB}(x-y) \cdot f_M(y + \bar{x}_M - \mu_0) dy,$$

da X und Y stochastisch unabhängig sind.

Sei $f_{NB}(x)$ die Verteilung der Variablen X im Kollektiv der "Nicht-Normalen", dann ist, bei Gültigkeit der gleichen Voraussetzungen,

$$[3] \quad f_N = f_{NB} * f_M = \int_{-\infty}^{+\infty} f_{NB}(x-y) \cdot f_M(y + \bar{x}_M - \mu_0) dy$$

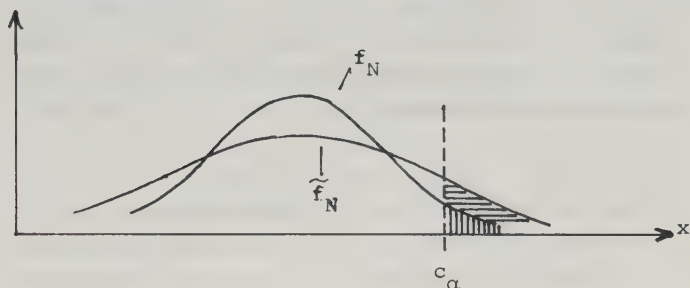
die Verteilung der Meßwerte der Variablen X im Kollektiv der Nicht-Normalen. Aus den Formeln [2] und [3] ist ersichtlich, daß bei einer Zunahme der Standardabweichung s_M des Meßfehlers die Verteilungen f_N und f_{NB} weiter auseinandergezogen und flacher werden und sich somit auch weiter überlappen, woraus eine Zunahme des Klassifikationsfehlers [1] resultiert (siehe Abb. 1.3.2).

Im weiteren Verlauf der Arbeit wird davon ausgegangen, daß der systematische Fehler $\bar{x}_M - \mu_0$ gleich Null ist.

Durch Einsetzen von [2] und [3] in [1] ergibt sich die Klassifikationsfehlerrate

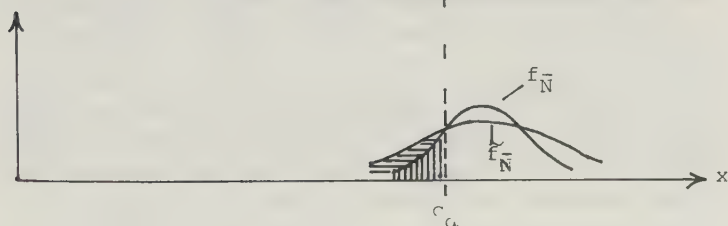
Abb. 1.3.2

Graphische Darstellung der Erhöhung der Rate von Falschklassifikationen von Patienten nach Erhöhung der Meßfehlervarianz von σ^2 auf $\tilde{\sigma}^2$.



f_N = Dichtefunktion des Normalpersonenkollektives mit Varianz σ_N^2

$\tilde{f}_N$ = Dichtefunktion des Normalpersonenkollektives nach Erhöhung der Meßfehlervarianz von σ_N^2 auf $\tilde{\sigma}_N^2$



$f_{\bar{N}}$ = Dichtefunktion des Kollektives der "Kranken" mit Varianz $\sigma_{\bar{N}}^2$

$\tilde{f}_{\bar{N}}$ = Dichtefunktion des Kollektives der "Kranken" nach Erhöhung der Meßfehlervarianz von $\sigma_{\bar{N}}^2$ auf $\tilde{\sigma}_{\bar{N}}^2$

 $\triangleq$ Anteil der falsch klassifizierten Meßwerte vor Erhöhung der Meßfehlervarianz.

 $\triangleq$ Anteil, um den die falsch klassifizierten Meßwerte nach Erhöhung der Meßfehlervarianz steigen.

$$[4] \quad q = p(N) \cdot \left(1 - \int_{c_1}^{c_2} f_{NB} * f_M(x) dx \right) + p(\bar{N}) \cdot \int_{c_1}^{c_2} f_{\bar{NB}} * f_M(x) dx$$

bei Diagnostik eines beliebigen (nicht präselektierten) Patienten aus der Grundgesamtheit.

1.4 Meßfehlervariabilität und "biologische Variabilität".

Gebräuchliche Methoden zur Bestimmung einer Toleranzgrenze für die Meßfehlervarianz.

Sei f_M die Meßfehlerverteilung mit Varianz σ_M^2 , sei f_{NB} die Verteilung einer Variablen des Normalpersonenkollektivs mit Varianz σ_{NB}^2 und sei $f_{\bar{NB}}$ die Verteilung derselben Variablen des "Nicht-Normalpersonenkollektivs" mit Varianz $\sigma_{\bar{NB}}^2$, dann ergeben sich bei Meßfehlerkonstanz die Verteilungen der Meßwerte dieser Variablen des Normalpersonenkollektivs als $f_N = f_{NB} * f_M$ und des "Nicht-Normalpersonenkollektivs" als $f_{\bar{N}} = f_{\bar{NB}} * f_M$ mit $\sigma_N^2 = \sigma_{NB}^2 + \sigma_M^2$ sowie $\sigma_{\bar{N}}^2 = \sigma_{\bar{NB}}^2 + \sigma_M^2$ (siehe Abb. 1.3).

Daraus folgt, daß eine Erhöhung der Meßfehlervarianz σ_M^2 um einen gewissen Prozentsatz eine umso größere Zunahme der Falschklassifikationen von Patientenproben durch den Normbereich bewirkt, je größer das Verhältnis von Meßfehlervarianz zur rein biologischen Varianz $\sigma_M^2 : \sigma_{NB}^2$ und $\sigma_M^2 : \sigma_{\bar{NB}}^2$ ist. Denn: Steigt die Meßfehlervarianz um y % an, dann ist die Varianz der Verteilung der Meßwerte der Variablen des Normalpersonenkollektivs $\sigma_N^2 = \sigma_{NB}^2 + ((y + 100) : 100) \cdot \sigma_M^2$ und die Varianz der Verteilung der Meßwerte derselben Variablen des "Nicht-Normalpersonenkollektivs" $\sigma_{\bar{N}}^2 = \sigma_{\bar{NB}}^2 + ((y + 100) : 100) \cdot \sigma_M^2$ (siehe 1.3). σ_N bzw. $\sigma_{\bar{N}}$ wird also umso größer, je größer σ_M relativ zu σ_{NB} bzw. $\sigma_{\bar{NB}}$ wird.

Es gibt mehrere Verfahren, die eine Toleranzgrenze für die Standardabweichung des zufälligen Meßfehlers s_M aus einer Relation zwischen s_M und der Standardabweichung s_N des Normalpersonenkollektivs ableiten (siehe 1.4.2). Anhand eines dieser Verfahren soll kurz dargestellt werden, daß bei diesen Verfahren eigentlich eine Relation zwischen Meßfehlervariabilität und biologischer Variabilität vorliegt und welcher Art diese Relation ist.

Eines dieser Verfahren wurde von SUCHET (37) publiziert.

Er postuliert:

Bezüglich einer bestimmten Stoffwechselvariablen darf die Standardabweichung s_M der Präzisionsstandardverteilung nicht größer sein als die Hälfte der Standardabweichung s_N der Meßwertverteilung des Normalpersonenkollektivs:

$$s_M \leq \frac{1}{2} s_N$$

Bei Meßfehlerkonstanz (siehe S. 8) ist die Varianz σ_N^2 der Meßwertverteilung des Normalpersonenkollektivs gleich der Summe der Varianzen der biologischen Variabilität des Normalpersonenkollektivs und der Meßfehlervariabilität:

$$\sigma_N^2 = \sigma_{NB}^2 + \sigma_M^2$$

Die Meßfehlervariabilität und die biologische Variabilität sind stochastisch unabhängig (siehe 1.3).

Aus SUCHET's Postulat $\sigma_M \leq \frac{1}{2} \sigma_N$ läßt sich dann ableiten:

$$\sigma_M^2 \leq \frac{1}{4} (\sigma_{NB}^2 + \sigma_M^2) \longrightarrow \sigma_M^2 - \frac{1}{4} \sigma_M^2 \leq \frac{1}{4} \cdot \sigma_{NB}^2$$

$$\longrightarrow \sigma_M \leq 0.577 \sigma_{NB}$$

Da s_N und s_M effiziente Schätzungen für σ_N und σ_M sind, ist SUCHET's (37) Postulat näherungsweise äquivalent mit der Bedingung, daß die Meßfehlervarianz nicht größer sein darf als die Hälfte der ("biologischen") Varianz der gemessenen Variablen im Normalpersonenkollektiv.

Diese Abschätzung ist nur bei Meßfehlerkonstanz auf dem Meßbereich korrekt. Meßfehlerkonstanz existiert zwar selten, die Unterschiede zwischen Meßfehlervarianzen in unterschiedlichen Meßbereichen (also bei unterschiedlichen Präzisionsstandardkonzentrationen) sind im allgemeinen jedoch so gering, daß obige Abschätzung beibehalten werden kann. SUCHET "erlaubt" dem Meßfehler relativ zur biologischen Variabilität also eine relativ große Schwankungsbreite. Daran mag man erkennen, wie schwer es offensichtlich ist, die Meßfehlervarianz um Größenordnungen kleiner zu halten, als die biologische Varianz.

In Deutschland gelten zur Beschränkung des zufälligen Meßfehlers die Richtlinien der Bundesärztekammer. Sie legen fest, daß der Variationskoeffizient $V = \frac{s_M}{\bar{x}_M} \cdot 100 \%$ der Präzisionsstandardverteilungen (auch relative Standardabweichung genannt) für eine Reihe von Substanzen einen gewissen Prozentsatz nicht übersteigen darf. Die Substanzen und die zugehörigen maximal zulässigen Variationskoeffizienten werden jährlich neu festgelegt und veröffentlicht.

Im Erhebungszeitraum der Präzisionsstandard-Meßreihen von 1971 bis 1974 galt z. B. ein maximal zulässiger Variationskoeffizient von 10 % für:

Eisen; Kupfer; Magnesium;
Phosphor, anorganisch; Bilirubin;
Harnstoff; Kreatinin; Kreatin;
Cholesterin; Phosphorlipide; Triglyzeride;
Fettfraktionen; Gesamt-Lipide;
Albumin; Gammaglobuline; Gerinnungsfaktor;
Immunglobuline; Aminosäuren;
Hormone und deren Metabolite;
Vitamine;
Leukozyten und Plättchen in elektronischen Zählgeräten;
Enzymaktivitätsbestimmungen

und für alle anderen Substanzen ein maximal zulässiger Variationskoeffizient von 5 %.

Leicht umgeformt lautet die Richtlinie der Bundesärztekammer also

$$s_M = \bar{x}_M \cdot 0.05 \text{ bzw. } s_M = \bar{x}_M \cdot 0.1.$$

Der "Pool" (siehe 1.5), dem die Präzisionsstandards entnommen werden, ist im allgemeinen ein Substanzgemisch von Serumproben von Patienten des klinisch-chemischen Labors, die in etwa dem Normalpersonenkollektiv entsprechen.

(Anmerkung: Im allgemeinen legt man bei der Ermittlung der Normbereiche für jede Stoffwechselvariable ein eigenes Normalpersonenkollektiv fest.)

Um einen relativ groben Vergleich zwischen SUCHET's Kriterium und

den Richtlinien der Bundesärztekammer machen zu können, mußten die Standardabweichungen der Stoffwechselvariablen des Normalpersonenkollektivs, die leider nicht explizit vorlagen, durch $\frac{1}{4} \cdot$ Normbereich approximiert werden; der Normbereich einer Stoffwechselvariablen umfaßt nämlich im allgemeinen das zwischen dem 2,5-Perzentil und dem 97,5-Perzentil liegende Intervall der empirischen Verteilung der Stoffwechselvariablen. Deren empirische Verteilungen im Normalpersonenkollektiv sind im allgemeinen Normalverteilungen ((36), (39), (44) aus (8)), das heißt, der Normbereich hat eine "Länge" von $4s_N$, das heißt $s_N = \frac{1}{4}$ Normbereich. SUCHET's Kriterium ist also annähernd der Bedingung $s_N = \frac{1}{8}$ Normbereich äquivalent.

In (8) wird als Kriterium der Begrenzung des zufälligen Meßfehlers die Bedingung

$$\frac{s_M}{\bar{x}_M} = \frac{1/4 \text{ Normbereich}}{\bar{x}_N}$$

angegeben, wobei $\bar{x}_M$ die Konzentration der gerade betrachteten Stoffwechselvariablen im Präzisionsstandard-Pool ist. Wenn diese nicht allzusehr vom Mittelwert $\bar{x}_N$ des Normalpersonenkollektivs dieser Stoffwechselvariablen abweicht, was auch im allgemeinen zu erwarten ist, dann ist dieses Kriterium ungefähr der Bedingung $s_M \leq \frac{1}{4}$ Normbereich äquivalent.

Das wohl schärfste Kriterium zur Begrenzung des zufälligen Meßfehlers wurde vom "College of American Pathologists" formuliert; es lautet: $s_M \leq \frac{1}{16}$ Normbereich.

SUCHET's Kriterium nimmt also eine Art Mittelstellung innerhalb

der Kriterien ein, die den maximal zulässigen zufälligen Meßfehler aus einer Relation zwischen empirischer Standardabweichung des Normalpersonenkollektives und der Standardabweichung des zufälligen Meßfehlers s_M gewinnen.

In Tab. 3.2.1 können anhand der dort aufgelisteten Größen $\frac{1}{8} \cdot$ Normbereich und $0.1 \cdot \bar{x}$ bzw. $0.05 \cdot \bar{x}_M$ (je nachdem, wie es die Richtlinie der Bundesärztekammer vorschreibt) die beiden Kriterien verglichen werden.

In dieser Tabelle kann ferner durch Vergleich der dort ebenfalls aufgeführten Standardabweichungen s_M der Präzisionsstandards nachgeprüft werden, ob die Bedingungen $s_M \leq \frac{1}{8} \cdot$ Normbereich bzw. $s_M \leq 0.05 \cdot \bar{x}_M$ bzw. $s_M \leq 0.1 \cdot \bar{x}_M$ erfüllt sind.

Die nach der Vorschrift der Bundesärztekammer errechneten Toleranzgrenze für die Meßfehlervarianz wird im folgenden mit dem Symbol s_{Mmax} bezeichnet.

1.5 Prinzip und Durchführung der Präzisionskontrolle

Die Methode, nach der im klinisch-chemischen Labor die Präzisionskontrolle durchgeführt wird, beruht auf der Annahme der Normalverteilung der Präzisionsstandards. Präzisionsstandards sind Proben aus einem Pool eines Substanzgemisches, das in etwa dem biologischen Grundstoff (z. B. Blutserum) eines Durchschnittsbürgers entspricht und in dem die Stoffwechselvariable (z. B. Kaliumkonzentration) überprüfbar ist, deren Meßfehler erfaßt werden soll. Durch Tiefgefrierung wird versucht, die Beschaffenheit dieses Pools über längere Zeit konstant zu halten. Nach Entnahme einer gewissen Anzahl (im allgemeinen 20) von Proben aus diesem Pool und nach Messung der Konzentration der interessierenden Stoff-

wechselvariablen in diesen Proben berechnet man den Mittelwert $\bar{x}_M$ und die Standardabweichung s_M dieser Meßwerte, welche in ihrer Gesamtheit auch Vorlauf genannt werden.

Wenn die Annahme der Normalverteilung der Präzisionsstandards zutrifft und wenn sich der Meßfehler nicht verändert, werden sich in Zukunft 99,7 % aller Präzisionsstandards im Intervall $(\bar{x}_M - 3s_M, \bar{x}_M + 3s_M)$ befinden, bei Erhöhung des Meßfehlers werden weniger, bei Erniedrigung mehr Präzisionsstandards in diesem Intervall liegen.

Zur praktischen Durchführung der Präzisionskontrolle werden Kontrollkarten benutzt (siehe Abb. 1.4.1). Falls ein Präzisionsstandard außerhalb der Kontrollgrenzen liegt, wird das Meßsystem als nicht mehr unter Kontrolle befindlich betrachtet.

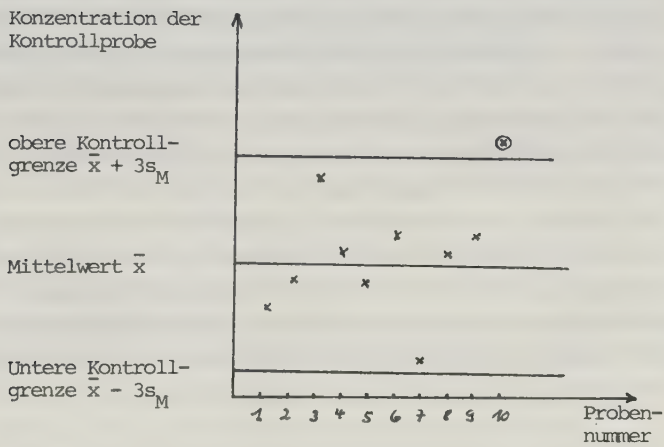
Sei in Abb. 1.1.2 f_{M1} die empirische Dichtefunktion des Vorlaufes und f_{M2} die empirische Dichtefunktion der Präzisionsstandards in einem Zeitintervall (t_1, t_2) nach Erhebung des Vorlaufes. Dann stellt die in Abb. 1.1.2 schraffierte Fläche den Prozentsatz von Präzisionsstandards dar, der in diesem Zeitintervall außerhalb der Kontrollgrenzen $\bar{x} - 3 \cdot s_{M1}$ und $\bar{x} + 3 \cdot s_{M1}$ zu liegen kommt.

Die Präzision des Meßsystems wird kontrolliert durch Beobachtung, ob ein Präzisionsstandard innerhalb der Grenzen $\bar{x}_M - 3s_M$ und $\bar{x} + 3s_M$, auch Kontrollgrenzen genannt (siehe Abb. 1.4.1), liegt oder nicht. In letzterem Falle wird das Meßsystem als nicht in Kontrolle befindlich betrachtet. Bei Zutreffen der oben gemachten Annahmen wird also in 0,3 % aller Fälle ein falscher Alarm gegeben.

Die Annahme, daß der Meßfehler nicht größer geworden ist aufgrund der Tatsache, daß ein einzelner Präzisionsstandard innerhalb der Kontrollgrenzen liegt, ist mit einer Fehlerwahrscheinlichkeit 0.34 behaftet (siehe Büttner (4), S. 344).

Abb. 1.4.1

Graphische Darstellung einer Kontrollkarte



2. Fragestellung

Alle klinisch-chemischen Labors sind vom Gesetzgeber verpflichtet, die Meßgenauigkeit ihrer chemischen Analyse-Verfahren ständig zu überwachen. Einer der dazu eingesetzten Kontrollmechanismen ist die in 1.5 beschriebene Präzisionskontrolle. Diese arbeitet nur dann korrekt, wenn die Zufallsvariable Meßfehler normalverteilt ist. Je mehr die Meßfehlerverteilung von einer Normalverteilung abweicht, desto fehlerhafter werden die Interpretationen der Ergebnisse der Präzisionskontrolle: Falls mehr als 0,3 % der Präzisionsstandards außerhalb des als Kontrollbereich festgelegten Intervalls ($\bar{x}_M - 3s_M$, $\bar{x}_M + 3s_M$) liegen, wird die Häufigkeit falscher Alarme ansteigen, im umgekehrten Falle wird die Verschlechterung der Präzision mit geringerer Wahrscheinlichkeit entdeckt, als aufgrund einer Normalverteilung der Präzisionsstandards zu erwarten wäre. In einem Meßsystem, dessen Meßfehler nicht normalverteilt ist, sind dann Fehlbeurteilungen der darin ermittelten Patientenergebnisse nicht mehr abschätzbar.

Die Überprüfung der Annahme, daß es sich bei der Verteilung der Präzisionsstandards um eine Normalverteilung handelt, ist also durchaus angezeigt und wichtig. Eine derartige Untersuchung ist bis jetzt nicht bekannt und soll Hauptgegenstand dieser Arbeit sein.

Die Prüfung einer Verteilung auf einen bestimmten Verteilungstyp hin erfolgt durch statistische Tests. Ein Test, der differenzierte Aussagen über Art und Umfang der Abweichung einer getesteten Verteilung von einer hypothetischen gestattet, ist der χ^2 -Anpassungstest, dessen Ergebnisse in erster Linie zur Bewertung der Präzisionsstandardverteilungen herangezogen wurden.

Ein in (42) publizierter Test auf Normalverteilung, der in erster Linie auf der Symmetrie der Normalverteilung aufbaut, erwies sich als weniger geeignet, insbesondere, da die zu testenden empirischen Verteilungen meist stark diskretisiert waren.

Wenn das Ergebnis des χ^2 -Anpassungstests eine Widerlegung der Annahme erbringt, daß es sich bei der untersuchten Meßreihe um eine Stichprobe aus einer normalverteilten Grundgesamtheit handelt, dann hat die Anwendung des allgemein üblichen Prinzips der Präzisionskontrolle ihre Berechtigung verloren. In diesem Falle stellt die Verwendung des 1-Perzentils und des 99-Perzentils als Kontrollgrenzen eine Alternative dar, deren Einsatzmöglichkeit in Kapitel 5.2 untersucht wird.

Zwar ist eine korrekte Kontrolle der Präzision nur bei Übereinstimmung der Präzisionsstandardverteilung mit einer Normalverteilung möglich, praktisch relevant sind Abweichungen der Präzisionsstandardverteilung von einer Normalverteilung aber nur dann, wenn die Standardabweichung s_M der Präzisionsstandardverteilung nahe an ihre Toleranzgrenze (siehe 1.4) heranreicht.

Deswegen werden in Tab. 3.2.1 für jede Präzisionsstandardmeßreihe deren Standardabweichung s_M sowie zwei nach verschiedenen Methoden ermittelte Toleranzgrenzen für s_M aufgeführt: Nämlich die aus SUCHET's (42) Kriterium abgeleitete obere Schranke $\frac{1}{8}$ Normbereich und die von der Bundesärztekammer vorgegebene obere Schranke s_{Mmax} , $s_{Mmax} = \bar{x} \cdot 0.05$ bzw. $s_{Mmax} = \bar{x} \cdot 0.1$, je nach klinisch-chemischen Analyseverfahren.

Die Frage, wie klein die Standardabweichung der Präzisionsstandards sein muß, damit auch bei nicht normalverteilten Präzisionsstandards die Effizienz der Präzisionskontrolle nicht unter ihr Minimum abfällt, wird in Kapitel 4.3 und 5.3 erörtert.

3. Beschreibung des Datenmaterials

3.1 Gewinnung der Daten

Es wurden Meßreihen von Präzisionsstandards der Jahre 1971 bis 1974 aus dem klinisch-chemischen Zentrallaboratorium der Friedrich-Alexander-Universität Erlangen-Nürnberg untersucht. Die zur Verfügung gestellten Meßreihen bestanden nicht nur aus den Werten der Vorläufe (siehe 1.4), sondern aus der Summe der Werte der Vorläufe und der Präzisionsstandards, die während der ganzen Charge als Kontrollproben mitgeführt wurden. Eine Charge ist die Menge der Proben, die in der Zeit analysiert werden, für die die aus dem Vorlauf ermittelten Kontrollgrenzen Gültigkeit besitzen.

Die zeitliche Begrenzung der Gültigkeit eines Vorlaufes und seiner Kontrollgrenzen ist aus zwei Gründen notwendig:

1. Der Meßfehler unterliegt nicht nur kurzzeitigen Intensitätsschwankungen von Stunde zu Stunde oder von Tag zu Tag, die es zu erkennen und zu beheben gilt, sondern auch längerfristigen Veränderungen von Präzision und Richtigkeit, die nicht immer korrigiert werden können. Einer solch veränderten "Großwetterlage" von Störeinflüssen muß das Kontrollsystem von Zeit zu Zeit angepaßt werden.
2. Der Pool der Präzisionsstandards ist nicht unbegrenzt haltbar. Er soll spätestens nach 6 Monaten durch einen neuen ersetzt werden. Die Verwendung von neuen Präzisionsstandards impliziert natürlich auch eine neue Bestimmung der Kontrollgrenzen.

Die einzelnen Verfahren zur Bestimmung der Stoffwechselvariablenkonzentration (siehe Tab. 3.2.1) wurden nach folgenden chemischen

Methoden durchgeführt, wobei bei den ersten 12 Verfahren das Gerät SMA 12/30 der Firma Technikon benutzt wurde.

Cholesterin/Serum Serie 1 - 5	Methode nach Liebermann- Burchardt
Calcium/Serum Serie 1 - 5	komplexometrisch
Phosphor anorg./Serum Serie 1 - 5	Phosphormolybdat
Bilirubin/Serum Serie 1 - 5	Jendrassik u. Grof
Albumin/Serum Serie 1 - 5	HABA-Reaktion
Eiweiß ges./Serum Serie 1 - 5	Biuret -Methode
Harnsäure/Serum Serie 1 - 5	Redox-Messung
Harnstoff-Stickst./Serum Serie 1 - 5	Diacetylmonoxim-Reaktion
Glucose/Serum Serie 1 - 5	Neocuproin-Methode
LDH/Serum Serie 1 - 5	Wacker-Methode
Alk. Phosphatase/Serum Serie 1 - 5	Methode nach King Armstrong
GOT/Serum Serie 1 - 5	Methode nach Karmen
Natrium/Serum	Flammen-Elektronen-spektro- skopisch
Kalium/Serum	Flammen-Elektronen-spektro- skopisch

Bicarbonat/Serum	titrimetrisch
Chlor/Serum	merkurimetrisch
Blutzucker	GOD/POD-Methode Eppendorf-Photometer
GOT/Serum Serie 6 - 7	kinetisch im UV-Test bei 25°
GPT/Serum Serie 6 - 7	kinetisch im UV-Test bei 25°
LDH/Serum Serie 6 - 7	kinetisch im UV-Test bei 25°
HBDH/Serum	kinetisch im UV-Test bei 25°
Cholesterin/Serum Serie 6	Methode nach Liebermann- Burchardt
Calcium/Serum Serie 6	komplexometrisch
Bilirubin/Serum Serie 6 - 7	Jendrassik u. Grof
Harnstoff-Stickst./Serum Serie 6 - 8	Diacetylmonoxim-Reaktion
Alk. Phosphatase/Serum Serie 6	P-NPP-Substrat-Messung

3.2 Empirische Verteilungen

Es wurden insgesamt 100 Präzisionsstandardmeßreihen untersucht. Da es den Rahmen dieser Arbeit sprengen würde, die Graphiken aller dieser empirischen Verteilungen hier darzustellen, werden in Abb. 3.2.1 und 3.2.2 zwei exemplarisch herausgegriffene empirische Verteilungen gezeigt. Um dem Leser dennoch einen Eindruck über die Qualität der untersuchten empirischen Verteilungen zu vermitteln, werden in der Tab. 3.2.1 folgende Charakteristika der Verteilungen aufgeführt: Stichprobenumfang n , Mittelwert $\bar{x}$, Standardabweichung s und der Quotient $v = \frac{n}{m}$, mit m = Anzahl der Massepunkte, auf die sich die beobachteten Werte verteilen. Dieser Quotient wurde deshalb in die Tabellen übernommen, da sich bei der Mehrzahl der untersuchten Meßreihen viele beobachtete Werte auf nur wenige Ausprägungen verteilten. Dieses, auch Diskretisierung genannte, Phänomen wird, nach Auskunft des klinisch-chemischen Zentrallaboratoriums der Universität Erlangen, in erster Linie durch mangelndes Auflösungsvermögen der Meßinstrumente verursacht. Es hat, wie in Kap. 4 noch ausführlicher beschrieben wird, Auswirkungen auf die Wahl der statistischen Tests und deren Durchführung. Von den meisten Stoffwechselvariablen lagen mehrere Meßreihen aus aufeinanderfolgenden Zeiträumen vor. Meßreihen, die in gleichen Zeiträumen erhoben wurden, sind in der Tab. 3.2.1 mit identischen Seriennummern gekennzeichnet.

Abb. 3.2.1

Empirische Dichtefunktion der Meßreihe von Serum-Calcium-Präzisionsstandards aus Serie 1

Stichprobenumfang: 124

Diskretisierter Definitionsbereich: $D(x) = 8.6 + 0.1 \cdot i$

$i = 0, \dots, 6$

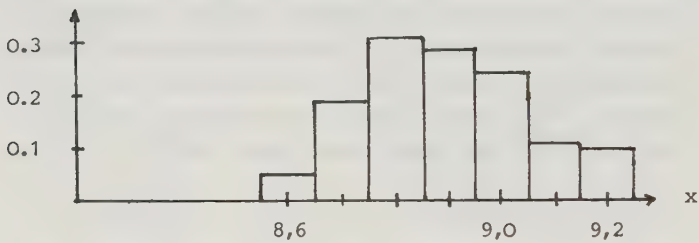
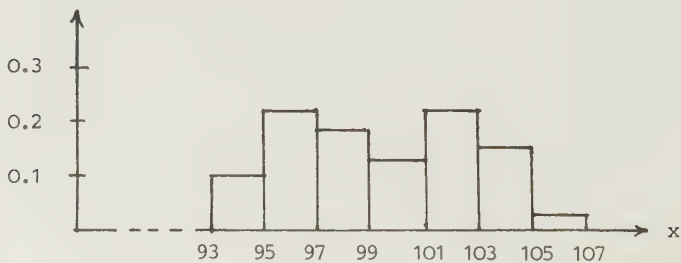


Abb. 3.2.2

Empirische Dichtefunktion der Meßreihe der Glucose-Präzisionsstandards aus Serie 3

Stichprobenumfang: 35

Anzahl von Ausprägungen: 20



Tab. 3.2.1

Kenngrößen der empirischen Verteilungen und obere Schranken für die Meßfehlervariabilität (Präzision)

Legende: Art : a: Meßwerteerhebung mit dem Laborautomaten
SMA 12 der Fa. Technicon.
m: "manuelle" Meßwerteerhebung mit Photo-
meter o. ä.

$\frac{1}{8}$ Normbereich : dient als Näherungswert für $\frac{1}{2} s_N$ (siehe 1.4)

s_{Mmax} : nach den Vorschriften der Bundesärztekammer
maximal zulässige obere Schranke für die
"Präzision" s_M .

v : bezeichnet das Verhältnis der Anzahl n der
beobachteten Werte einer Meßreihe zur Anzahl
m der Ausprägungen (Massepunkte).

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert x	Stan- dard- abwei- chung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
anorgan. Phosphor	1	a	126	3.39	0.090	0.249	0.34	25.0:1
	3	a	52	4.00	0.117	0.249	0.40	4.5:1
	4	a	39	4.16	0.060	0.249	0.42	2.6:1
	5	a	130	4.09	0.110	0.249	0.41	11.9:1

Fortführung

Tab. 3.2.1

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert $\bar{x}$	Stan- dard- abwei- chung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
Chol- esterin	1	a	119	173.4	6.25	18.7	17.3	13.6:1
	2	a	34	166.6	3.77	18.7	16.7	2.6:1
	3	a	52	160.5	5.89	18.7	16.1	4.0:1
	4	a	39	167.9	5.39	18.7	16.8	2.6:1
	5	a	117	165.0	6.77	18.7	16.5	7.1:1
	6	m	65	151.9	6.90	22.5	15.2	2.9:1
Calcium	1	a	124	8.90	0.160	0.249	0.445	13.8:1
	2	a	106	4.86	0.118	0.249	0.123	13.3:1
	3	a	52	4.74	0.070	0.249	0.137	2.6:1
	4	a	39	4.78	0.080	0.249	0.139	1.7:1
	5	a	130	4.71	0.080	0.249	0.136	26.0:1
	6	m	39	4.92	0.125	0.124	0.146	6.5:1
Bili- rubin	1	a	127	0.51	0.040	0.112	0.050	63.0:1
	3	a	52	2.25	0.133	0.112	0.225	2.9:1
	4	a	39	2.32	0.105	0.112	0.232	1.9:1
	5	a	130	2.28	0.129	0.112	0.228	21.6:1
	6	m	39	4.45	0.058	0.112	0.445	2.0:1
	7	m	39	1.84	0.082	0.112	0.184	1.7:1

Fortführung

Tab. 3.2.1

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert $\bar{x}$	Stan- dard- abwei- chung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
Albumin	1	a	126	4.16	0.070	0.183	0.416	31.6:1
	2	a	76	4.31	0.107	0.183	0.431	15.0:1
	3	a	52	4.51	0.083	0.183	0.451	3.5:1
	4	a	39	4.43	0.056	0.183	0.443	1.9:1
	5	a	130	4.40	0.091	0.183	0.440	26.0:1
Eiweiß	1	a	126	6.08	0.105	0.249	0.304	15.0:1
	3	a	52	7.23	0.206	0.249	0.362	2.2:1
	4	a	39	7.39	0.087	0.249	0.365	1.3:1
	5	a	130	Ausreißer				13.0:1
Harnsäure	1	a	120	5.020	0.228	0.687	0.251	11.0:1
	3	a	52	5.131	0.446	0.687	0.261	1.5:1
	4	a	39	5.600	0.230	0.687	0.280	1.5:1
	5	a	130	5.000	0.450	0.687	0.250	6.2:1

Fortführung

Tab. 3.2.1

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert $\bar{x}$	Stan- dard- abweichung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
Harnst. Stickst.	1	a	95	15.88	0.790	1.250	0.794	19.0:1
	3	a	52	16.00	1.040	1.250	0.800	3.0:1
	4	a	39	16.65	0.290	1.250	0.833	3.3:1
	5	a	130	15.98	0.810	1.250	0.799	11.8:1
	6	m	52	15.40	0.980	1.250	0.770	13.0:1
	7	m	65	16.20	0.780	1.250	0.810	13.0:1
	8	m	95	4.90	0.120	1.250	0.222	19.0:1
Glucose	1	a	120	84.16	6.613	5.62	4.208	15.0:1
	3	a	52	89.87	17.100	5.62	4.499	2.0:1
	4	a	39	99.60	3.740	5.62	4.980	1.9:1
	5	a	130	100.50	7.826	5.62	5.030	4.5:1
Serum-LDH	1	a	116	71.20	6.24	11.25	7.12	19.3:1
	3	a	52	81.16	5.64	11.25	8.12	2.6:1
	4	a	39	88.80	3.24	11.25	8.88	1.3:1
	5	a	130	89.60	6.24	11.25	8.96	4.5:1
	6	m	65	463.10	36.80	13.13	46.31	5.9:1
	7	m	39	465.50	37.53	13.13	46.55	3.5:1

Fortführung

Tab. 3.2.1

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert $\bar{x}$	Stan- dard- abwei- chung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
alkal. Phospha- tase	1	a	124	7.34	0.510	1.62	0.73	44.6:1
	2	a	60	26.70	1.550	1.62	2.67	12.0:1
	3	a	52	18.00	1.030	1.62	1.80	3.0:1
	4	a	39	19.70	0.670	1.62	1.57	2.4:1
	6	a	52	51.40	4.600	3.75	5.14	4.7:1
Serum-GOT	1	a	109	22.21	4.11	3.75	2.22	15.5:1
	3	a	52	32.33	5.12	3.75	3.23	2.6:1
	4	a	39	34.00	2.40	3.75	3.40	1.7:1
	5	a	130	32.60	5.18	3.75	3.26	5.4:1
	6	m	65	38.37	2.42	16.50	3.84	5.4:1
	7	m	39	40.47	1.90	16.50	4.05	4.3:1
Natrium	1	m	128	142.3	1.65	2.86	7.12	12.8:1
	2	m	39	140.4	1.39	2.86	7.02	6.5:1
	3	m	78	141.9	1.73	2.86	7.10	9.8:1
Kalium	1	m	229	4.02	0.06	0.170	0.20	32.3:1
	2	m	39	4.29	0.09	0.170	0.21	4.9:1
	3	m	78	4.33	0.08	0.170	0.22	15.6:1

Fortführung

Tab. 3.2.1

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert $\bar{x}$	Stan- dard- abwei- chung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
Chlorid	1	m	125	101.6	1.02	1.00	5.08	25.0:1
	2	m	39	101.3	1.17	1.00	5.07	6.5:1
	3	m	78	102.4	1.33	1.00	5.12	13.0:1
Blut- zucker	1	m	208	83.5	6.10	5.0	4.18	7.7:1
	2	m	260	88.3	3.71	5.0	4.42	13.0:1
	3	m	65	85.0	2.70	5.0	4.25	5.7:1
Bicarbonat		m	39	12.1	1.09	0.75	0.62	6.5:1
Kreatinin		m	107	0.89	0.08	0.0225	0.089	21.0:1
Serum-GPT	1	m	65	44.3	2.09	2.74	2.22	7.2:1
	2	m	39	42.5	2.76	2.74	2.13	3.5:1
HBDH		m	65	259.4	15.429	15.0	25.94	2.2:1
		m	39	250.5	19.540	15.0	25.05	2.4:1
saure Phosphatase		m	65	22.77	0.930	1.375	2.277	2.8:1
		m	39	26.36	0.503	1.375	2.636	3.0:1

Fortführung

Tab. 3.2.1

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert $\bar{x}$	Stan- dard- abwei- chung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
Serum- Amylase		m	65	1866	230.126	309	186	1.3:1
		m	52	2158	276.066	309	215	1.2:1
Choleste- rinase		m	39	1511	142.802	237.5	151.1	2.0:1
			30	-	-	237.5	-	5.0:1
Gesamt- lipide		m	65	518.2	29.451	62.5	51.82	4.6:1
		m	78	-	-	62.5	-	7.8:1
α_1 -Globu- line		m	91	5.558	0.679	0.625	0.558	2.8:1
α_2 -Globu- line		m	104	9.837	0.804	0.750	0.9837	2.7:1
β -Globuline		m	91	11.79	0.985	0.725	1.179	2.3:1
γ -Globuline		m	91	16.12	1.254	1.16	1.612	2.1:1
Harn-Natrium		m	26	80.00	2.184	10.6	4.00	3.7:1
		m	52	139.50	1.394	10.6	6.98	8.6:1

Fortführung

Tab. 3.2.1

Stoff- wechsel- variablen	Se- rie	Art	An- zahl n	Mittel- wert $\bar{x}$	Stan- dard- abwei- chung s_M	$\frac{1}{8}$ Norm- bereich	s_{Mmax}	v
Harn-Kalium	m		26	27.54	1.255	3.12	1.377	3.2:1
	m		52	Ausreißer		3.12		7.2:1
Harn-Cal- cium	m		26	4.448	0.213	1.88	0.2224	2.9:1
	m		36	4.851	0.096	1.88	0.2426	4.3:1

4. Test der empirischen Präzisionsstandardverteilungen auf Normalverteilung

4.1 Zur Auswahl eines Tests auf Normalverteilung

Gebräuchliche Verfahren zur Beurteilung der Übereinstimmung einer empirischen Verteilung mit einer Normalverteilung sind der Kolmogorov-Smirnov-Test, ein nicht näher bezeichneter Test auf Normalverteilung nach VAN DER WAERDEN, und der χ^2 -Anpassungstest.

Obwohl beim Kolmogorov-Smirnov-Test die bei diskretisierten empirischen Verteilungen oft etwas problematische Einteilung der Abszisse in Intervalle unterbleibt, wurde er hier nicht in die engere Auswahl genommen - aus zwei Gründen:

1. Beim Kolmogorov-Smirnov-Test geht in die Testgröße nur die maximale Abweichung zwischen empirischer Verteilung und hypothetischer Verteilung ein, während beim χ^2 -Anpassungstest die Abweichung zwischen empirischer und hypothetischer Verteilung aller Intervalle der Abszisse eingeht.
2. Der χ^2 -Anpassungstest liefert, im Gegensatz zum Kolmogorov-Smirnov-Test, als Zwischenergebnis die prozentualen Abweichungen der empirischen Verteilung von der hypothetischen in jedem Intervall, welche für den "Präzisionskontrollleur" des klinisch-chemischen Labors durchaus von Interesse sein können.

4.1.1 Ein nicht näher bezeichneter Test nach VAN DER WAERDEN
(42), Seite 229)

Seien $m_k = \frac{1}{n} \cdot \sum (x - \bar{x})^k$ die zentralen Momente der empirischen Verteilung, dann sind

$$g_1 = m_3 \cdot m_2^{-\frac{3}{2}} \quad \text{und} \quad g_2 = m_4 \cdot m_2^{-2} - 3 \text{ Schätzwerte}$$

für die Schiefe $\gamma_1 = \mu_3 / \mu_2^{-\frac{3}{2}}$ und den Exzess $\gamma_2 = \mu_4 / \mu_2^{-2} - 3$

Für Stichproben mit endlichem Umfang n ersetze man

$$\text{die Schiefe } g_1 \text{ durch } G_1 = \frac{n(n-1)}{n-2} g_1 \quad \text{und}$$

$$\text{den Exzess } g_2 \text{ durch } G_2 = \frac{(n-1)(n+1)}{(n-2)(n-3)} (g_2 + 6)$$

Unter der Annahme der Normalverteilung sind die Erwartungswerte von G_1 und G_2 exakt 0. Ihre Varianzen sind

$$\sigma_1^2 = \frac{6n(n-1)}{(n-2)(n+1)(n+3)} \quad \sigma_2^2 = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}$$

Man kann also G_1 / σ_1 oder G_2 / σ_2 als Testgröße für Normalität verwenden. Beide Größen sind asymptotisch normalverteilt mit Mittelwert 0 und Streuung 1. Das heißt: Befindet sich G_1 / σ_1 oder G_2 / σ_2 nicht innerhalb des Intervalls $[\bar{x} - 2s, \bar{x} + 2s]$, also im Intervall $[-2, +2]$, dann ist die Hypothese, daß es sich bei der getesteten Verteilung um eine Normalverteilung handelt, mit einer Irrtumswahrscheinlichkeit von ca. 4,5 % abzulehnen.

Da die in diesen Test eingehenden Größen die einzelnen Meßwerte selbst und nicht die Häufigkeiten dieser Werte in den einzelnen In-

tervallen sind, dürfen die aus der Anwendung dieses Tests auf diskretisierte empirische Verteilungen resultierenden Ergebnisse nur mit Vorsicht interpretiert werden - aus zwei Gründen:

Nehmen wir an, die Grundgesamtheit habe eine glockenförmige Dichtefunktion und der Mittelwert befinde sich sehr nahe am Scheitelpunkt dieser Funktion - was für Präzisionsstandardverteilungen zutrifft. Nehmen wir weiter an, die Meßinstrumente haben ein nur endliches Auflösungsvermögen. Die dadurch verursachte Diskretisierung ist einer Histogrammbildung äquivalent, bei der die einzelnen Stichprobenwerte zu den Ausprägungen (ablesbare Skalenteile) hin auf- bzw. abgerundet werden, zu denen sie den geringsten Abstand haben. Die Intervallgrenzen des so gebildeten Histogrammes entsprechen also jeweils den Mittelpunkten zwischen zwei Ausprägungen.

Bei Annahme einer glockenförmigen Dichtefunktion ist vor Diskretisierung die Dichte an den Intervallrändern größer, die näher am Scheitelpunkt der Dichtefunktion liegen. Durch die Abbildung aller Werte eines Intervalls auf seinen Mittelpunkt wird die Summe der Abstände der Stichprobenwerte vom Mittelwert

$$\sum_{i=1}^n (x_i - \bar{x})$$

vergrößert. Die Summe dieser einzelnen Abstände geht aber gerade nach einer Potenzierung in die zentralen Momente m_2 und m_3 ein, die zur Berechnung der Testgröße aus Normalverteilung dienen. Andererseits hätten aber ohne Diskretisierung die weiter vom Scheitelpunkt entfernten, weniger häufigen Werte einen höheren Beitrag zu den zentralen Momenten geliefert, da deren Abstände $(x_i - \bar{x})$ mit der 2. bzw. 3. Potenz in die Berechnung eingehen.

Darüber, wie weit sich diese beiden Fehler unter Umständen ausgleichen, war in der Literatur keine Information erhältlich. Die eben angestellten Überlegungen legen allerdings die Annahme nahe, daß die durch Diskretisierung verursachten Fehler bei der Berechnung der zentralen Momente stark von der Gestalt der empirischen Dichtefunktion der betrachteten Meßreihe abhängen.

Eine Abschätzung dieser Fehler wäre Voraussetzung für eine mit kalkulierbarem Fehlerrisiko ausgeführte Interpretation der Ergebnisse dieses Tests.

Da diese Abschätzung nur bei Kenntnis der nichtdiskretisierten empirischen Verteilung möglich ist, und da diese Voraussetzung nicht gegeben ist, wird auf einen Beitrag des Ergebnisses dieses Tests zur Entscheidung, ob die Hypothese der Normalverteilung angenommen wird oder nicht, verzichtet.

Bei 17 % der empirischen Verteilungen führten die Ergebnisse des eben beschriebenen Tests und des χ^2 -Anpassungstests zu konträren Entscheidungen. Da dies ausschließlich stark diskretisierte Verteilungen waren, wie der Leser durch Vergleich der Testergebnisse in Tab. 4.2.1 mit dem Quotienten v in Tab. 3.2.1 ersehen kann, wurden diese Diskrepanzen in erster Linie der eben beschriebenen Fehlerquelle zugeordnet.

4.1.2 Der χ^2 -Anpassungstest

Das Prinzip des χ^2 -Anpassungstests besteht darin, über die empirische Verteilung $F_E(x)$ und die hypothetische Verteilung $F_H(x)$ ein identisches Intervallraster zu legen, in jedem Intervall die "Abweichung" der empirischen Verteilung von der in der Hypothese zugrunde gelegten zu messen, die Einzelabweichungen zum Ergebnis χ_o^2 aufzusummieren und dann durch Vergleich von χ_o^2 mit einer "kritischen Zahl c "

zu entscheiden, ob die Hypothese, daß die empirische Verteilung $F_E(x)$ mit der Verteilung $F_H(x)$ "übereinstimmt", zu verwerfen ist oder nicht. Folgendes "Rezept" des χ^2 -Anpassungstests wurde aus KREYSZIG (21), Seite 230, entnommen:

Tabelle 84.1. Chi-Quadrat-Test für die Hypothese, daß $F(x)$ die Verteilungsfunktion der Grundgesamtheit ist, der die Stichprobe $x_1, \dots, x_n$ entstammt

1. *Schritt.* Man unterteile die x -Achse in K Intervalle $I_1, I_2, \dots, I_K$, und zwar derart, daß jedes Intervall mindestens 5 Werte der gegebenen Stichprobe $x_1, \dots, x_n$ enthält. Für jedes Intervall I_j bestimme man die Anzahl b_j der Stichprobenwerte, die in I_j liegen. Ein Wert, der auf eine Intervallgrenze fällt, wird je zur Hälfte in den beiden angrenzenden Intervallen gezählt.

2. *Schritt.* Aus $F(x)$ berechne man für jedes Intervall I_j die Wahrscheinlichkeit p_j , mit der die betreffende Zufallsvariable X irgendeinen Wert aus I_j annimmt. Daraus berechne man die Anzahl

$$e_j = n p_j$$

der theoretisch in I_j zu erwartenden Stichprobenwerte.

3. *Schritt.* Man berechne die Abweichung

$$(84.1) \quad \chi_0^2 = \sum_{j=1}^K \frac{(b_j - e_j)^2}{e_j}.$$

4. *Schritt.* Man wähle eine Signifikanzzahl α (5%, 1% oder dgl.).

5. *Schritt.* Man bestimme die Lösung c der Gleichung

$$P(\chi^2 \leq c) = 1 - \alpha$$

aus der Tafel 6 der Chi-Quadrat-Verteilung mit $K - 1$ Freiheitsgraden. Ist $\chi_0^2 \leq c$, so wird die Hypothese angenommen. Ist $\chi_0^2 > c$, so wird sie verworfen.

Zum 5. Schritt. Enthält die zu testende Verteilungsfunktion r Parameter, deren genaue Werte nicht bekannt sind, so bestimme man zuerst Näherungen für diese Parameter mit Hilfe der Maximum-Likelihood-Methode (siehe Abschn. 67) und benutze dann die Chi-Quadrat-Verteilung mit $K - r - 1$ (statt $K - 1$) Freiheitsgraden.

Beim χ^2 -Anpassungstest auf Normalverteilung, wie im vorliegenden Fall, ist $F_H(x)$ die Normalverteilung mit Mittelwert μ und Varianz σ^2 : $F_H(x) = N(\mu, \sigma^2)$, wobei man μ durch $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ und σ^2 durch $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ abschätzt, also $F_H(x)$ durch $N(\bar{x}, s)$ ersetzt.

Sei $I_j = (h_j, j_{j+1})$ das j -te Intervall, dann errechnet sich die laut Hypothese in diesem Intervall erwartete Zahl e_j von Werten als

$$e_j = n \cdot \left[\Phi\left(\frac{h_{j+1} - \bar{x}}{s}\right) - \Phi\left(\frac{h_j - \bar{x}}{s}\right) \right]$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung $N(0,1)$ ist.

KREYSZIG (21) gibt im gleichen Kapitel noch an, daß jeweils so viele Intervalle zusammengefaßt werden sollen, bis $e_j = 5$.

Die Intervallbildung wurde nach folgenden verschiedenen Methoden ausgeführt:

4.1.2.1 "Gängige Methode" (VAN DER WAERDEN (42))

Die Spanne zwischen dem größten und dem kleinsten Wert (Spannweite) wird in m äquidistante Intervalle eingeteilt. m ergibt sich aus

$\sqrt{n}$, aufgerundet auf die nächste ganze Zahl, wobei n der Stichprobenumfang ist.

4.1.2.2 Intervallgrenze als Mittelpunkt zwischen benachbarten Ausprägungen

Wie schon in 3. beschrieben, waren mehr als 50 % der Meßreihen stark diskretisiert, wobei die Abstände zwischen den Ausprägungen durchwegs konstant waren. Bei diesen Meßreihen kann man davon ausgehen, daß die Intervallbildung schon infolge mangelnden Auflösungsvermögens der Meßinstrumente sowie durch Schätzung und Interpolation der angezeigten Werte auf bestimmte Punkte der Meßskala vorgenommen wurde. In diesen Fällen ist es nicht sinnvoll, durch erneute Intervallbildung artifizielle Leerklassen zu provozieren, da diese das Testergebnis überproportional beeinflussen und verfälschen können. Hier wurden als Intervallgrenzen jeweils die Mittelpunkte zwischen benachbarten Ausprägungen gewählt.

4.2 Procedere bei der Durchführung des χ^2 -Anpassungstests mit Hinweisen auf das ausführende Fortran-Programm

Die empirischen Verteilungen wurden unter Einsatz eines im Rahmen dieser Arbeit entwickelten Fortran-Programmes dem χ^2 -Anpassungstest unterzogen, wobei jede Verteilung nach den beiden in 4.1.2.1 und 4.1.2.2 beschriebenen Methoden der Intervallbildung getestet wurde.

Bei annähernd kontinuierlichen empirischen Präzisionsstandardverteilungen dienten die Ergebnisse der nach der in 4.1.2.1 beschriebenen Variante durchgeführten Tests als Grundlage der Entscheidung, ob die Hypothese der Normalverteilung der Präzisionsstandards verworfen wird oder nicht. Bei diskreten empirischen Präzisionsstandardverteilungen dienten die Ergebnisse der nach der in 4.1.2.2 beschriebenen Variante durchgeführten Tests als Bewertungsmaßstab. Die dis-

kretisierten empirischen Verteilungen sind am hohen Quotienten v in Tab. 3.2.1 erkennbar.

Anwender des Programmes, das den χ^2 -Anpassungstest ausführt, finden die im Intervall i laut Hypothese erwartete Anzahl von Werten e_i in der mit $E(I)$ bezeichneten Spalte, die Anzahl der tatsächlich beobachteten Werte in der Spalte " $B(I)$ ", das Abweichungsquadrat im Intervall i in der Spalte " AB ". Das Ergebnis χ^2_0 ist dort als " ABS " gekennzeichnet.

Die zur Berechnung der Größen $E(I)$ und ABS benötigten Wertepaare der Standard-Normalverteilung liegen dem Programm als Datenkarten bei. Die Schrittweite auf der Abszisse beträgt 0.01.

Wie schon in 4.1.2 gezeigt, ist die "Anpassung" einer empirischen Verteilung an eine Normalverteilung umso besser, je kleiner ihr

χ^2_0 -Wert unter Berücksichtigung der Zahl der Freiheitsgrade ist, das heißt, je kleiner das Quantil der χ^2 -Verteilung ist, unterhalb dessen gerade noch der χ^2_0 -Wert der empirischen Verteilung liegt. Die gängigen Tabellen von χ^2 -Werten in Abhängigkeit von der Zahl der Freiheitsgrade umfassen die Quantile:

0.001	;	0.005	;	0.01	;	0.02	;	0.05	;	0.1
0.25	;	0.50	;	0.75	;	0.90	;	0.975	;	0.99
0.995	;	0.999	;							

Für die Freiheitsgrade von 1 bis 50 sind die χ^2 -Werte der oben genannten Quantile dem Programm ebenfalls auf Datenkarten beigelegt, so daß das Testergebnis direkt - ohne Nachschlagen in Tabellen - der vom Computer ausgedruckten Liste entnommen werden kann.

Bei Zugrundelegung einer Signifikanzzahl von $\alpha = 5\%$ wird bei einem Testergebnis $F(\chi_o^2) > 0.95$ die Hypothese, die Präzisionsstandards seien normalverteilt, abgelehnt, bei einem Wert von $F(\chi_o^2) < 0.95$ angenommen.

Abb. 4.4.1 bis Abb. 4.4.4 veranschaulichen einen Schritt bei der Durchführung des χ^2 -Anpassungstests:

In jedem auf der Abszisse markierten Intervall sind zwei Säulen eingetragen. Die "blanke" Säule symbolisiert die Anzahl der in dem Intervall beobachteten Werte und die schraffierte die laut Hypothese erwartete Anzahl von Werten.

In Tab. 4.4.1 stehen in der Spalte "FG" die Freiheitsgrade (aus der empirischen Verteilung berechnet) und rechts daneben in der Spalte "c" die kritischen Werte c, für die die χ^2 -Verteilung $F_{\chi^2, FG}$ mit FG Freiheitsgraden den Wert $F_{\chi^2, FG}(c) = 0.95$ annimmt. Die Freiheitsgrade errechnen sich aus der Zahl der Intervalle K, in die die empirische Verteilung eingeteilt wird, und der Anzahl r der geschätzten Parameter der hypothetischen Verteilung (bei der Normalverteilung μ und σ): $FG = K - r - 1$ (bei der Normalverteilung $FG = K - 3$).

Die Hypothese, die Stichprobe entstamme einer normalverteilten Grundgesamtheit, gilt als angenommen, wenn $\chi_o^2 \leq c$. Die χ_o^2 -Werte der empirischen Verteilung stehen in der gleichnamigen Spalte.

In der Spalte "Variablen" stehen die untersuchten Stoffwechselvariablen.

Von den meisten Variablen standen mehrere Meßreihen zur Verfügung. Die Spalte "Serie" enthält die Nummer, die der zeitlichen Reihenfolge der Meßserie entspricht. Für die Beurteilung der Ergebnisse ist dabei wichtig, daß Meßreihen verschiedener Stoffwechselvariablen mit gleicher Seriennummer dem gleichen Zeitintervall entstammen, das heißt deren Präzisionsstandards wurden dem gleichen Pool entnommen.

4.3 Einführung einer weitgehend verteilungsunabhängigen Toleranzgrenze für die Präzision

Falls die Hypothese, die Präzisionsstandardverteilung sei eine Normalverteilung, abgelehnt wird, so besteht die Möglichkeit, daß - bedingt durch die Abweichung der Präzisionsstandardverteilung von einer Normalverteilung - eine Verschlechterung der Präzision entweder mit größerer oder mit kleinerer Wahrscheinlichkeit entdeckt wird, als dies bei Normalverteilung der Präzisionsstandards der Fall wäre.

Im folgenden wird versucht, eine Toleranzgrenze für die Präzision zu finden, bei deren Einhaltung eine Verschlechterung der Präzision auch bei nicht normalverteilten Präzisionsstandards mindestens mit gleicher Wahrscheinlichkeit entdeckt wird, wie bei deren Normalverteilung.

Zur Bestimmung dieser "verteilungsfreien" Toleranzgrenze wird die Annahme gemacht, daß die Präzisionsstandardverteilungen nicht mehrgipfelig sind. Theoretisch wird diese Annahme dadurch begründet, daß mehrgipfelige Präzisionsstandardverteilungen nur dann auftreten, wenn sich der systematische Fehler (siehe Kap. 1.2) sprunghaft verändert und daß solche Veränderungen durch die Richtigkeitskontrolle entdeckt werden.

Die 100 untersuchten empirischen Verteilungen widersprachen dieser Annahme im großen und ganzen nicht: 8 der 100 untersuchten empirischen Präzisionsstandardverteilungen waren zweigipfelig, wobei allerdings anzumerken ist, daß die Wahl der Intervalle bei der Ermittlung der empirischen Verteilungen einen nicht unerheblichen Einfluß auf die Anzahl der "Gipfel" hat.

Unter der Annahme, daß die Präzisionsstandardverteilungen nicht mehrgipfelig sind, ist die Wahrscheinlichkeit, mit der eine Verschlechterung der Präzision erkannt wird, dann am geringsten, wenn die Präzisionsstandards gleichverteilt sind. Die Standardabweichung kann dann um den Faktor 1.73 zunehmen, die Präzision sich also um den gleichen Faktor verschlechtern, ohne daß ein Meßwert außerhalb des Kontrollbereichs liegt.

Begründung: Bei einer Gleichverteilung gilt zwischen Standardabweichung s und Spannweite S folgende Beziehung: $S = \sqrt{12} \cdot s = 3.46 s$

$$\begin{aligned} \text{Beweis: } s^2 &= \int_{-\infty}^{+\infty} (\bar{x} - x)^2 f(x) dx = \int_{-\frac{1}{2}S}^{+\frac{1}{2}S} (0 - x)^2 \frac{1}{S} dx \\ &= \frac{2}{S} \int_0^{\frac{1}{2}S} x^2 dx = \frac{1}{12} S^2 \end{aligned}$$

Innerhalb des Intervalles $(\bar{x} - 1.73s, \bar{x} + 1.73s)$ liegen dann 100 % des Fläche der Verteilung, während die Kontrollgrenzen das Intervall $(\bar{x} - 3s, \bar{x} + 3s)$ umfassen.

Die Standardabweichung kann also um den Faktor p , $p = \frac{2 \cdot 3}{2 \cdot 1.73} = 1.73$ zunehmen (aus $2 \cdot 3s = 2 \cdot 1.73s \cdot p$), ohne daß ein Präzisionsstandard außerhalb der Kontrollgrenzen zu liegen kommt.

Falls also die Standardabweichung der Präzisionsstandardverteilung

$$s_M < (1 : 1.73) \cdot s_{Mmax} \quad (s_{Mmax} \text{ siehe Kap. 1.4}),$$

wird durch das übliche Verfahren der Präzisionskontrolle, auch wenn die Präzisionsstandardverteilung keine Normalverteilung ist, garantiert, daß eine Überschreitung von s_{Mmax} durch s_M mindestens mit gleicher Wahrscheinlichkeit wie bei der Normalverteilung der Präzisionsstandards entdeckt wird.

In Tab. 4.3.1 sind in der mit "Präzision" überschriebenen Spalte die Präzisionsstandardverteilungen, für die

$$s_M \leq (1 : 1.73) \cdot s_{Mmax}$$

gilt, mit "++" gekennzeichnet, diejenigen, für die

$$(1 : 1.73) \cdot s_{Mmax} < s_M \leq s_{Mmax}$$

gilt, mit "+" und diejenigen, für die

$$s_{Mmax} < s_M$$

gilt, mit "-".

In Kapitel 5.2 wird noch gezeigt, daß bei $s_M \leq (1 : 1.73) \cdot s_{Mmax}$ auch bei bis auf Eingipfeligkeit beliebiger "Form" der Präzisionsstandardverteilung die maximal zulässige Rate φ_{max} von Falschklassifikationen von Patientenmeßwerten durch den Normbereich schwerlich überschritten werden kann, das heißt bei den mit "++" gekennzeichneten Präzisionsstandardverteilungen ist eine Ablehnung der Hypothese der Normalverteilung eigentlich ohne Bedeutung.

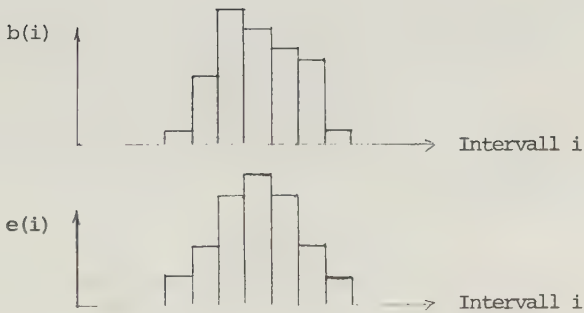
Abb. 4.2.1

Ergebnis des χ^2 -Anpassungstests angewandt auf die empirische Verteilung der Serum-Calcium-Präzisionsstandards der ersten Serie.

Legende:

$b(i)$ = Anzahl von beobachteten Werten im Intervall I

$e(i)$ = laut Hypothese erwartete Anzahl von Werten im Intervall i



$$\chi^2_o = 4.687$$

$$c = 9.49$$

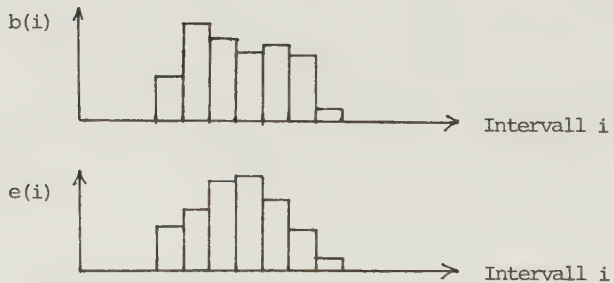
Abb. 4.2.2

Ergebnis des χ^2 -Anpassungstests angewandt auf die Serum-Glucose-Präzisionsstandards der zweiten Serie

Legende:

$b(i)$ = Anzahl von beobachteten Werten im Intervall i

$e(i)$ = laut Hypothese erwartete Anzahl von Werten im Intervall i



$$\chi^2_o = 5.763$$

$$c = 9.49$$

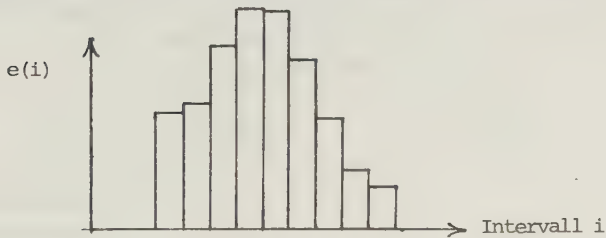
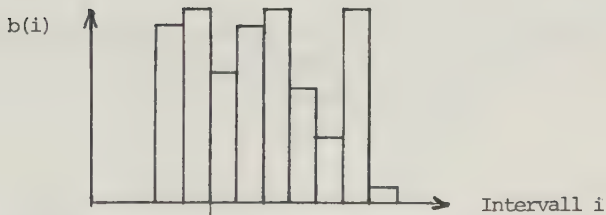
Abb. 4.2.3

Ergebnis des χ^2 -Anpassungstests angewandt auf die empirische Verteilung der Serum-Gesamtlipide.

Legende:

$b(i)$ = Anzahl von beobachteten Werten im Intervall i

$e(i)$ = laut Hypothese erwartete Anzahl von Werten im Intervall i



$$\chi^2_{0} = 27.656$$

$$c = 12.59$$

Tab. 4.4.1

Legende:

Die Spalte bezeichnet mit

enthält folgende Werte

n	Anzahl der beobachteten Werte einer Meßreihe
FG	Anzahl der Freiheitsgrade (siehe 4.1.2)
c	kritischer Wert, für den die χ^2 -Verteilung mit FG Freiheitsgraden den Wert $F \chi^2_{FG(c)} = 0.95$ annimmt
$G_1 / \tilde{G}_1$	Testgröße auf Normalverteilung nach VAN DER WAERDEN (42): falls $-2 < G_1 / \tilde{G}_1 < +2$, wird die Hypothese, die Meßreihe entstamme einer normalverteilten Grundgesamtheit, nicht verworfen.
Testergebnis	"-" bei Ablehnung der Hypothese mit Irrtumswahrscheinlichkeit 5 % nach χ^2 -Anpassungstest, "+" sonst
Präzision	"++", falls $s_M \leq (1 : 1.73) \cdot s_{Mmax}$ "+", falls $(1 : 1.73) \cdot s_{Mmax} < s_M \leq s_{Mmax}$ "-", falls $s_M > s_{Mmax}$ (siehe Kap. 5.2)

Fortführung siehe
nächste Seite

Stoff- wechsel- variab- len	Se- rie	Art	An- zahl n	$G_1/5_1$	FG	χ^2	c	Test- ergeb- nis	"Prä- zi- sion"
anorg. Phosphor	1	a	126	1.9	2	1.20	5.99	+	++
	3	a	52	0.8	1	5.89	3.84	-	++
	4	a	39	-2.1	1	1.15	3.84	+	++
	5	a	130	0.4	3	23.66	7.81	-	++
Choleste- rin	1	a	119	-1.8	3	2.80	7.81	+	++
	2	a	34	-0.9	3	7.70	7.81	+	++
	3	a	52	-0.9	2	4.50	5.99	+	++
	4	a	39	-3.2	1	3.77	3.84	+	++
	5	a	117	1.5	10	33.66	16.92	-	++
	6	m	65	-2.2	6	10.20	12.59	+	++
Calcium	1	a	124	2.80	4	4.60	9.49	+	++
	2	a	106	0.14	2	5.10	5.99	+	+
	3	a	52	2.40	2	2.20	5.99	+	++
	4	a	39	-0.70	4	5.00	9.49	+	++
	5	a	130	-0.40	1	0.90	3.84	+	+
	6	m	39	0.76	1	0.28	3.84	+	+

Fortführung

Tab. 4.4.1

Stoff- wechsel- variab- le	Se- rie	Art	An- zahl n	G_1/σ_1	FG	χ^2_{α}	c	Test- ergeb- nis	"Prä- zi- sion"
Biliru- bin	1	a	127	-6.00	0	-	-	-	+
	3	a	52	-0.70	3	2.40	7.81	+	+
	4	a	39	1.80	1	0.95	3.84	+	++
	5	a	130	0.59	2	6.14	5.99	-	++
	6	m	39	1.04	1	0.56	3.84	+	++
	7	m	39	-1.70	2	1.66	5.99	+	++
Albumin	1	a	126	0.77	1	2.10	3.84	+	++
	2	a	76	0.84	2	4.35	5.99	+	++
	3	a	52	-2.90	3	24.50	7.81	-	++
	4	a	39	-0.40	2	0.40	5.99	+	++
	5	a	130	0.50	2	3.17	5.99	+	++
Eiweiß	1	a	76	0.10	2	2.73	5.99	+	++
	3	a	52	-1.60	5	7.50	11.07	+	++
	4	a	39	-0.44	2	2.97	5.99	+	++
	5	a	130	20.00	Ausreißer		-	.	

Fortführung

Tab. 4.4.1

Stoff- wechsel- variab- len	Se- rie	Art	An- zahl n	G_1/b_1	FG	χ^2	c	Test- ergeb- nis	"Prä- zi- sion"
Harnsäure	1	a	110	0.11	5	7.10	11.07	+	+
	3	a	52	-2.10	5	13.40	11.07	-	-
	4	a	39	0.87	2	2.00	5.99	+	-
	5	a	130	2.30	6	5.75	12.59	+	-
Harnst. Stickst.	1	a	95	-0.18	0	-	-	+	+
	3	a	52	0.80	5	23.60	11.07	-	-
	4	a	39	0.20	2	0.70	5.99	+	++
	5	a	130	-0.80	5	34.50	11.07	-	-
	6	m	52	4.70	1	3.74	3.84	+	-
	7	m	65	-0.62	0	-	-	+	+
	8	m	39	0.70	1	0.13	5.99	+	++
Glucose	1	a	120	-1.50	3	13.80	7.81	-	-
	3	a	52	-5.00	1	4.30	3.84	-	-
	4	a	39	0.50	3	3.24	7.81	+	+
	5	a	130	0.12	7	35.30	14.07	-	-

Fortführung

Tab. 4.4.1

Stoff- wechsel- variab- len	Se- rie	Art	An- zahl n	$G_1/\bar{G}_1$	FG	χ^2_0	c	Test- ergeb- nis	"Prä- zi- sion"
Natrium	1	m	120	1.27	4	8.76	9.49	+	++
	2	m	39	-0.35	3	3.50	7.81	+	++
	3	m	78	-1.00	4	6.40	7.81	+	++
Kalium	1	m	129	2.23	0	4.88	-	-	++
	2	m	39	0.28	2	29.00	5.99	-	++
	3	m	78	0.05	1	2.00	3.84	+	++
Chlorid	1	m	125	1.04	2	0.854	5.99	+	++
	2	m	39	1.50	1	1.539	3.84	+	++
	3	m	78	0.77	2	17.820	5.99	-	++
Blut- zucker	1	m	208	-3.000	10	44.48	18.31	-	-
	2	m	260	-0.250	10	46.08	18.31	-	+
	4	m	65	-1.027	5	4.70	11.07	+	++
Bicarbo- nat		m	39	0.70	3	0.56	7.81	+	-
Kreatinin		m	107	-0.245	2	52.29	5.99	-	+

Fortführung

Tab. 4.4.1

Stoff- wechsel- variab- len	Se- rie	Art	An- zahl n	G_1/σ_1	FG	χ^2 c	Test- ergeb- nis	"Prä- zi- sion"
Serum-GPT		m	65	-0.06	4	1.845 9.49	+	+
		m	39	-0.20	1	1.547 3.84	+	-
HBDH		m	65	2.10	4	5.474 9.49	+	+
		m	39	1.10	2	13.550 5.99	-	+
Saure Phospha- tase		m	65	0.20	2	9.25 5.99	-	++
		m	39	2.50	3	0.83 7.81	+	++
Serum- Amylase		m	65	1.20	8	9.55 15.51	+	-
		m	52	1.80	4	5.49 9.49	+	-
Choleste- rinase		m	39	4.20	1	5.525 3.84	-	+
Gesamt- lipide		m	65	-2.20	6	2.42 12.59	+	++
		m	78	-	6	27.66 12.59	-	
α_1 -Globuline		m	91	-0.90	5	4.29 11.07	+	-
α_2 -Globuline		m	104	-0.30	6	9.23 12.59	+	+

Fortführung

Tab. 4.4.1

Stoff- wechsel- variab- len	Se- rie	Art	An- zahl n	$G_1/\bar{G}_1$	FG	χ^2 σ^2	Test- ergeb- nis	"Prä- zi- sion"
β-Globuline		m	91	-0.50	5	1.041 11.07	+	+
γ-Globuline		m	91	-1.00	4	6.243 9.49	+	+
Harn-Natrium		m	26	1.10	5	5.947 11.07	+	++
			52	1.00	1	4.882 3.84	+	++
Harn-Kalium		m	26	0.001	1	4.681 3.84	-	+
			52	15	Ausreißer			
Harn-Calcium		m	26	1.00	1	0.818 3.84	+	+
			36	-1.40	2	6.717 5.99	-	++

5. Diskussion

5.1 Ergebnisse und deren Interpretation

Um zu prüfen, ob die in der klinischen Chemie allgemein gemachte Annahme, daß die Meßwerte von Präzisionskontrollproben normalverteilt sind, zutrifft, wurden in dieser Arbeit insgesamt 99 Meßreihen von Präzisionskontrollproben 31 verschiedener Stoffwechselvariablen auf Normalverteilung getestet.

Drei Meßreihen enthielten Ausreißer und konnten somit nicht ausgewertet werden. Der Umfang der Meßreihen variierte zwischen 39 und 260 beobachteten Werten und betrug im Mittel ca. 85. Von den wichtigsten Stoffwechselvariablen standen jeweils vier Meßreihen zur Verfügung. Es handelt sich dabei um mit dem Laborautomaten SMA-12 der Fa. Technicon erhobene Meßreihen folgender Stoffwechselvariablen (gemessen im Serum):

Cholesterin, Calcium, anorg. Phosphor, Gesamt-Bilirubin, Albumin, Gesamt-Eiweiß, Harnsäure, Harnstoff-Stickstoff, Glucose, LDH, alk. Phosphatase, GOT, in ihrer Gesamtheit auch als S-Tests bezeichnet. Von Cholesterin, Calcium, Bilirubin, Harnstoff-Stickstoff, LDH, alk. Phosphatase, GOT und Glucose lagen außerdem noch Meßreihen vor, die mit anderen Laborgeräten (Eppendorf-Photometer) erhoben worden waren.

Die empirischen Verteilungen, die im χ^2 -Anpassungstest 0 Freiheitsgrade hatten, wurden nach dem von VAN DER WAERDEN vorgeschlagenen Test (siehe 4.1.1) beurteilt. Bei 32 der 96 auswertbaren Meßreihen (also bei ca. 33 %) wurde die Hypothese der Normalverteilung mit einer Irrtumswahrscheinlichkeit von 5 % abgelehnt. Gehäuft war dies bei folgenden Stoffwechselvariablen der Fall:

Liste 1: Anorg. Phosphor im Serum	bei 2 von 4 Meßreihen
Bilirubin im Serum	bei 2 von 6 Meßreihen
Harnstoff-Stickstoff im Serum	bei 2 von 7 Meßreihen
Glucose im Serum	bei 5 von 7 Meßreihen
LDH im Serum	bei 3 von 6 Meßreihen
Alk. Phosphatase im Serum	bei 2 von 5 Meßreihen
GOT im Serum	bei 3 von 6 Meßreihen
Kalium im Serum	bei 2 von 3 Meßreihen

Die Annahme, die Präzisionsstandards seien normalverteilt, trifft also offensichtlich nicht so generell zu, wie allgemein angenommen wird.

Ziel der Präzisionskontrolle ist es, die beim Meßvorgang und in dessen Vorfeld auftretenden zufälligen Fehler zu begrenzen und somit zu einer Limitierung der Klassifikationsfehlerrate φ von Patientenmeßwerten durch den Normbereich (siehe Kap. 1.3, Gleichung [4]) beizutragen.

Die Klassifikationsfehlerrate φ wird aber mehr von der Größe der zufälligen Meßfehler (= Präzision) als von der Form der Meßfehlerverteilung (= Präzisionsstandardverteilung) beeinflusst. Dem wird im Rahmen der Präzisionskontrolle dadurch Rechnung getragen, daß geprüft wird, ob die Standardabweichung der Präzisionsstandards des Vorlaufes (siehe Kap. 1.4 und 1.5) unterhalb der vorgeschriebenen Toleranzgrenze s_{Mmax} liegt (siehe Kap. 1.4).

Bei Anwendung der von der Bundesärztekammer vorgeschriebenen Toleranzgrenze wurde bei 18 der insgesamt 96 auswertbaren Präzisionsstandardverteilungen, also bei ca. 19 %, die Standardabweichung als zu hoch befunden. Die betreffenden Stoffwechselvariablen sind in Tab. 4.4.1 mit "-" gekennzeichnet. Gehäuft kam dies bei folgenden Stoffwechselvariablen vor:

Liste 2: Harnsäure im Serum	bei 3 von 4 Meßreihen
Harnstoff-Stickstoff im Serum	bei 3 von 7 Meßreihen
Glucose im Serum	bei 4 von 7 Meßreihen
GOT im Serum	bei 3 von 6 Meßreihen
Amylase im Serum	bei 2 von 2 Meßreihen

Vergleicht man Liste 1 mit Liste 2, so fällt auf, daß es bei den Stoffwechselvariablen Harnstoff-Stickstoff, Glucose und GOT im Serum offensichtlich schwierig ist, die zufälligen Meßfehler quantitativ (die Standardabweichung s_M überschreitet die Toleranzgrenze s_{Mmax}) und qualitativ (Hypothese der Normalverteilung wird abgelehnt) in den Griff zu bekommen.

Bei der Untermenge der Meßreihen, bei denen $s_M > s_{Mmax}$, liegt der Prozentsatz von Ablehnungen der Normalverteilung bei 56 %, während er in der Gesamtmenge aller Meßreihen bei 33 % liegt.

In Tab. 3.2.1 fällt auf, daß die Variabilität des Stichprobenumfanges in der Gesamtmenge aller empirischen Verteilungen groß, innerhalb der durch gleiche Seriennummern gekennzeichneten Untermengen jedoch klein ist. Dies wird vermutlich durch verschieden lange Beobachtungszeiträume von Meßwerten bzw. durch verschieden hohe Beobachtungsfrequenzen von Meßwerten in den einzelnen Serien verursacht. Einer unterschiedlich langen Beobachtungsdauer in den einzelnen Serien wird auch die Serienabhängigkeit der Testergebnisse bei den 12 Stoffwechselvariablen des S-Tests zugeschrieben: Bei den Meßreihen der vierten Serie war in allen Fällen die Hypothese der Normalverteilung anzunehmen, die empirischen Verteilungen waren überwiegend gut an Normalverteilungen angepaßt, wie man den niedrigen χ_o^2 -Quadratwerten in Tab. 4.4.1 entnehmen kann. Es handelt sich dabei um Meßreihen niedrigen Umfanges ($n = 39$), also um Meßreihen,

die wahrscheinlich in einem kurzen Zeitraum erhoben wurden. Da Meßfehlerveränderungen umso wahrscheinlicher auftreten, je länger der Beobachtungszeitraum ist, und da diese im allgemeinen eine Abweichung der empirischen Präzisionsstandardverteilungen von einer Normalverteilung bewirken (siehe Kap. 1.3), wird als Ursache der Serienabhängigkeit der Testergebnisse eine unterschiedliche Länge der Meßreihenbeobachtungszeiträume der einzelnen Serien angenommen.

5.2 Verfahren bei Ablehnung der Hypothese der Normalverteilung der Präzisionsstandards

Bei Ablehnung der Hypothese, daß die Präzisionsstandards einer Variablen normalverteilt sind, verbleiben folgende Verfahrensweisen, die Präzision des Meßsystems dieser Variablen zu kontrollieren:

- 1) Man findet und eliminiert die Ursachen, die die Abweichung der Präzisionsstandardverteilung von der Normalverteilung verursachen.
- 2) Man ermittelt die "verteilungsfreien" Perzentile der empirischen Verteilungen der Präzisionsstandards dieser Variablen und verwendet diese als Kontrollgrenzen.
- 3) Man findet einen anderen Verteilungstyp, dem die Präzisionsstandardverteilungen dieser Variablen besser angepaßt sind als dem der Normalverteilung und verwendet Perzentile dieser Verteilung als Kontrollgrenzen.

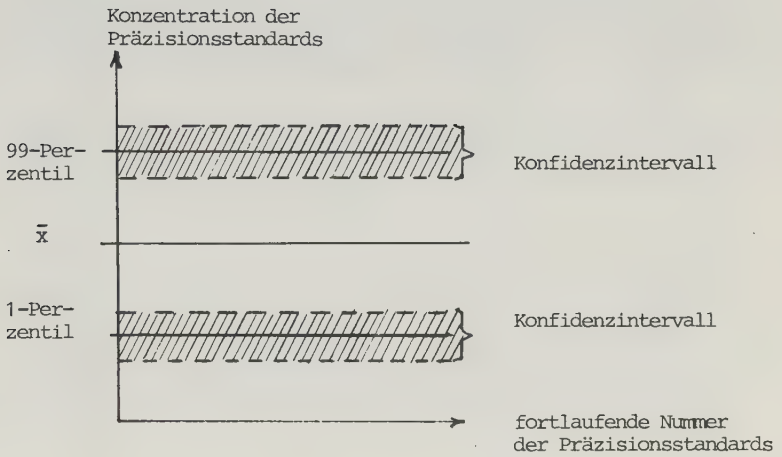
- Ad 1) Das erste Verfahren ist das erstrebenswerteste, da, bei gelungener Elimination der die Abweichung der Präzisionsstandardverteilung von einer Normalverteilung bewirkenden Ursachen, mit großer Sicherheit alle nicht zufälligen Fehlerquellen beseitigt sind, und somit die Voraussetzungen für die Gültigkeit des Modells der Präzisionskontrolle wieder gegeben sind.
- Ad 2) Falls dies nicht gelingt, besteht die Möglichkeit, verteilungsfreie Perzentile einzuführen. Dies sind Perzentile, die aus einer empirischen Verteilung berechnet werden, ohne daß die Verteilung der Grundgesamtheit bekannt ist. Die Ablehnung der Hypothese der Normalverteilung der Präzisionsstandards ist ein Zeichen dafür, daß es nicht möglich ist, den zufälligen Fehler des Meßsystems separat zu erfassen. Dies und die Unkenntnis der Verteilung der Grundgesamtheit der Präzisionsstandards machen kalkulierbare Aussagen über den Zustand des Meßsystems kaum möglich. Aus diesem Grunde verbietet sich der simple Ersatz der Kontrollgrenzen $\bar{x}_M - 3s_M$ und $\bar{x}_M + 3s_M$ durch das 1-Perzentil bzw. 99-Perzentil. Vielmehr sollten dann die Konfidenzintervalle dieser beiden Perzentile als Rückweisungsklassen eingeführt werden: das heißt, daß keine Entscheidung über die Präzision des Meßsystems gefällt wird, wenn ein Präzisionsstandard in ein Konfidenzintervall zu liegen kommt (siehe Abb. 5.1). Die verteilungsfreien Perzentile einer empirischen Verteilung nehmen aber einen umso größeren Anteil an deren Spannweite ein, je geringer der Stichprobenumfang und je größer die Anzahl deren gebundener Werte ist. Gebundene Werte sind Werte, die auf die gleiche Ausprägung "fallen".

Die Berechnung obengenannter Konfidenzintervalle bei einigen ausgesuchten empirischen Präzisionsstandardverteilungen, die die geringste Anzahl gebundener Werte und den größten Stichprobenumfang aufwiesen, ergab, daß die Wahrscheinlichkeit dafür, daß ein Präzisionsstandard in das Konfidenzintervall des 5-Perzentils oder des 95-Perzentils zu liegen kommt, mindestens 20 % beträgt. Bei den diskretisierten empirischen Verteilungen bzw. bei den Verteilungen kleineren Stichprobenumfanges liegt dieser Prozentsatz noch höher. Er beträgt bis zu 60 %. Daraus geht die Unpraktikabilität dieses Verfahrens zur Präzisionskontrolle der hier untersuchten Variablen hervor. Tab. 5.1 zeigt einen Arbeitsgang bei der Ermittlung der obengenannten verteilungsfreien Perzentile der empirischen Verteilung der Glucose-Präzisionsstandards in der 3. Serie (siehe Tab. 3.2.1). Nach dem in den wissenschaftlichen Tabellen von GEIGY beschriebenen Verfahren wurde für das 5-Perzentil Q (0.05) das Konfidenzintervall $(-\infty, 75.0)$ und für das 95-Perzentil das Konfidenzintervall $(102.0, +\infty)$ mit einer Irrtumswahrscheinlichkeit von 5 % ermittelt. Damit liegen 58 % der Spannweite und 32 % der Werte der empirischen Verteilung innerhalb dieser Konfidenzintervalle. Das heißt, bei jeder dritten Präzisionsstandardmessung wird eine Doppelbestimmung oder eine andere Meßnahme zur Herbeiführung einer definitiven Aussage über den Zustand des Meßsystems nötig.

- Ad 3) Dieses Verfahren ist, ebenfalls wie das unter 2) beschriebene, nur dann sinnvoll anzuwenden, wenn sich der zufällige Meßfehler mit keinen Mitteln separat erfassen läßt und wenn genügend empirische Verteilungen von Präzisionsstandards derselben Variablen vorhanden sind, so daß nachgewiesen werden kann, daß die empirischen Präzisionsstandardverteilungen mit der hypo-

thetischen - mit einer Normalverteilung nicht identischen - Verteilung über einen längeren Zeitraum hin übereinstimmen. Es werden dann viele Verteilungen derselben Variablen benötigt, um einen Sicherheitsbereich dafür angeben zu können, daß die hypothetische Verteilung mit den empirischen Verteilungen der betreffenden Variablen in Zukunft übereinstimmt.

Abb. 5.1



Tab. 5.1

Tabelle zur Bestimmung von verteilungsfreien Konfidenzintervallen

Ordnungs- zahl o	Wert x	Häufigkeit $f(x) = f(o)$	Summenhäufigkeit $F(x) = F(o)$
1	63.8	1	1
2	65.0	1	2
3	70.0	1	3
4	73.9	2	5
5	75.0	3	8
6	75.1	1	9
7	77.7	1	10
8	78.3	1	11
9	80.0	3	14
10	84.6	1	15
11	85.0	5	20
12	89.7	1	21
13	90.0	4	25
14	95.0	4	29
15	96.0	3	32
16	97.9	3	35
17	99.1	1	36
18	100.0	6	42
19	102.0	2	44
20	105.0	1	45
21	110.0	1	46
22	114.0	1	47
23	115.0	3	50

5.3 Beurteilung der Testergebnisse der empirischen Präzisionsstandardverteilungen auf Normalverteilung unter Berücksichtigung der Präzision

Die Gütekontrollen im klinisch-chemischen Labor, wie z. B. die Richtigkeitskontrolle und die Präzisionskontrolle, dienen durch das Erkennen - und somit eventuell durch die Beschränkung - einer Meßfehlerzunahme eigentlich der Limitierung der Klassifikationsfehlerrate $\mathcal{G}$ von Patientermesswerten (siehe Kap. 1.3).

Explizit werden in den Vorschriften der Bundesärztekammer für die einzelnen Stoffwechselvariablen keine maximal zulässigen Klassifikationsfehlerraten $\mathcal{G}_{\max}$ angegeben.

Implizit wird jedoch durch die Vorgabe einer Toleranzgrenze für die Präzision (= Präzisionsstandardabweichung) - unter der Annahme, daß die Präzisionsstandards normalverteilt sind - eine obere Schranke $\mathcal{G}_{\max}$ definiert: Bei Normalverteilung der Präzisionsstandards (der systematische Fehler wird in dieser Arbeit nicht betrachtet) ergibt sich $\mathcal{G}_{\max}$ aus der Fehlerrate $\mathcal{G}$, wenn die Präzision s_M an der (oberen) Toleranzgrenze $s_{M\max}$ liegt (siehe Kap. 1.4).

In Kap. 4.3 wurde eine weitgehend "verteilungsfreie" Toleranzgrenze für die Präzision eingeführt, bei deren Anwendung in der Präzisionskontrolle eine Verschlechterung der Präzision auch bei beliebiger Präzisionsstandardverteilung mit mindestens gleicher Wahrscheinlichkeit entdeckt wird, wie bei Normalverteilung der Präzisionsstandards. Falls es sich nachweisen ließe, daß die maximal zulässige Klassifikationsfehlerrate $\mathcal{G}_{\max}$ nicht überschritten wird, wenn die Präzisionsstandards zwar nicht normalverteilt sind, deren Standardabweichung (= Präzision) aber unter der "verteilungsfreien Toleranzgrenze" bleibt, dann wäre eine Ablehnung der Hypothese, die Präzisionsstan-

dards seien normalverteilt, ohne Relevanz für die Präzisionskontrolle.

Exakt läßt sich dieser Nachweis jedoch nicht mit vertretbarem Aufwand führen; die Klassifikationsfehlerrate φ müßte unter allen möglichen Verteilungsformen einer Stoffwechselvariablen sowohl des Kollektivs der "Normalen" als auch des Kollektivs der "Nicht-Normalen" sowie unter allen Verteilungsformen der Präzisionsstandards betrachtet werden.

Es läßt sich jedoch abschätzen, um wieviel der Wert, den die Standardabweichung der Meßergebnisse einer Stoffwechselvariablen des Normalpersonenkollektivs bei Anwendung der "verteilungsfreien" Toleranzgrenze für die Präzision maximal erreichen kann, unter dem Wert liegt, den diese Standardabweichung unter den üblichen Bedingungen der Präzisionskontrolle maximal erreichen kann. Unter Verwendung der von der Bundesärztekammer vorgeschriebenen Toleranzgrenze s_{max} für die Präzision läßt sich diese Abschätzung nicht durchführen, da diese Toleranzgrenze nicht für alle Stoffwechselvariablen durch dieselbe Funktion darstellbar ist. Deshalb wird für diese Abschätzung die von SUCHET (37) vorgeschlagene Toleranzgrenze $\frac{1}{2} \cdot s_N = 0.577 s_{\text{NB}}$ verwandt, die gut mit den von der Bundesärztekammer vorgeschriebenen Toleranzgrenzen übereinstimmt (siehe Kap. 1.4 und Tab. 3.2.1), wobei s_N = Standardabweichung der Meßwerte der Stoffwechselvariablen im Normalpersonenkollektiv und s_{NB} = ("biologische") Standardabweichung der Stoffwechselvariablen im Normalpersonenkollektiv.

Da der zufällige Meßfehler und die Stoffwechselvariablenkonzentration statistisch unabhängige Zufallsvariable sind, gilt bei Meßfehlerkonstanz (siehe Kap. 1.1) der Additionssatz für Varianzen. Sei s_{N1} die unter den üblichen Bedingungen der Präzisionskontrolle maximal zulässige Standardabweichung des Normalpersonenkollektivs, dann gilt

$$s_{N1}^2 = s_{NB}^2 + (0.577 s_{NB})^2 = 1.334 s_{NB}^2 \quad (\text{siehe Kap. 1.4}).$$

Sei s_{N2} der maximal zulässige Wert für die Standardabweichung des Normalpersonenkollektivs, der bei Anwendung der "verteilungsfreien" Toleranzgrenze für die Präzision erreicht werden darf, dann gilt:

$$s_{N2}^2 = s_{NB}^2 + (1 : 1.73) \cdot 0.577 \cdot s_{NB}^2 = 1.111 s_{NB}^2$$

Aus $s_{N2} : s_{N1} = 1 : 0.855$ folgt, daß die maximale Streuung einer Stoffwechselvariablen im Normalpersonenkollektiv bei Anwendung der "verteilungsfreien" Toleranzgrenze für die Präzision um den Faktor 0.855 kleiner ist als die maximal erreichbare Streuung des Normalpersonenkollektiv unter den üblichen Bedingungen der Präzisionskontrolle.

Da die untersuchten Präzisionsstandardverteilungen keine besonders pathologischen Formen aufwiesen, dürfte diese Erniedrigung der maximal zulässigen Streuung einer Stoffwechselvariablen im Normalpersonenkollektiv um den Faktor 0.855 ausreichen, um zu gewährleisten, daß die maximal zulässige Klassifikationsfehlerrate g_{max} auch dann nicht überschritten wird, wenn zwar die Annahme der Normalverteilung der Präzisionsstandards nicht zutrifft, die Präzision aber unterhalb der "verteilungsfreien" Toleranzgrenze liegt.

Bei 10 von 32 Meßreihen, bei denen die Hypothese der Normalverteilung abgelehnt wurde, lag die Standardabweichung unter dieser "verteilungsfreien" Toleranzgrenze. Es sind dies jene empirischen Verteilungen, die in Tab. 4.4.1 in der Spalte "Testergebnis" ein "-" und in der Spalte "Präzision" ein "++" aufweisen.

Das heißt, nur bei 21 der in 5.1 erwähnten 32 Meßreihen kann infolge der Abweichung von einer Normalverteilung mit Überschreitung der

maximal zulässigen Klassifikationsfehlerrate gerechnet werden. (Bei einer der 32 Meßreihen, bei denen die Hypothese der Normalverteilung abgelehnt wurde, konnte die Präzision nicht ermittelt werden.)

Diese Meßreihen weisen in Tab. 4.4.1 in Spalte "Testergebnis" ein "-" und in Spalte "Präzision" ein "+" oder ein "-" auf. Gehäuft wurde dies bei folgenden Stoffwechselvariablen beobachtet:

Liste 3: Glucose im Serum	bei 5 von 7 Meßreihen
LDH im Serum	bei 3 von 6 Meßreihen
alk. Phosphatase im Serum	bei 2 von 5 Meßreihen
GOT im Serum	bei 3 von 6 Meßreihen

Zusammenfassend läßt sich sagen, daß bei einem - auch für Labormediziner erstaunlich hohen - Anteil von $\frac{1}{3}$ der untersuchten Präzisionsstandard-Meßreihen die Hypothese der Normalverteilung abgelehnt werden mußte. Auch nach Ausfiltern derjenigen Meßreihen, bei denen aufgrund der niedrigen Standardabweichung nicht mit einem Überschreiten der maximal zulässigen Klassifikationsfehlerrate von Patientenmeßwerten zu rechnen ist, treffen noch bei ca. $\frac{1}{5}$ aller Meßreihen die bezüglich der Meßgenauigkeit gemachten Annahmen nicht zu.

6. Zusammenfassung

In dieser Arbeit wurden 99 empirische Präzisionsstandard-Meßreihen 31 verschiedener Stoffwechselvariablen untersucht. Drei Meßreihen enthielten Ausreißer und wurden deshalb nicht ausgewertet. Bei 18 empirischen Präzisionsstandard-Meßreihen überschritt die Standardabweichung die von der Bundesärztekammer vorgeschriebene Toleranzgrenze.

Die Präzisionsstandard-Meßreihen wurden auf Normalverteilung geprüft. Hierzu wurde ein Fortran-IV-Programm entwickelt, das die Verteilungen der Meßreihen berechnet und den χ^2 -Anpassungstest für eine jeweils geeignete Klasseneinteilung durchführt. Bei 32 der 96 auswertbaren Präzisionsstandard-Meßreihen wurde die Hypothese der Normalverteilung bei einem Signifikanzniveau von 5 % abgelehnt.

Weiterhin wurde ein Kriterium entwickelt, das es gestattet, auch bei nicht normalverteilter Meßfehlerverteilung die Klassifikationsfehlerrate von Patientenmeßwerten abzuschätzen. 10 der insgesamt 32 Präzisionsstandardverteilungen, bei denen die Hypothese der Normalverteilung abgelehnt wurde, erfüllten dieses Kriterium. In 22 Fällen konnte nicht ausgeschlossen werden, daß die Präzisionskontrolle die ihr zuge dachte Funktion nicht erfüllt hat. Gehäuft kam dies bei den Präzisionsstandardverteilungen der Stoffwechselvariablen Harnstoff-Stickstoff, Glucose, LDH, alk. Phosphatase und GOT vor.

7. Literaturverzeichnis

1 Adlassnig, K. P., et al.:

A Computer Assisted System for Diagnostic Decision Making -
Online Usage of the Database of the Medical Information System
WAMIS.

In Shires, D. B., Wolff, H. (Eds.): MEDINFO 77, North-Holland,
Amsterdam (1977)

2 Bechtler, G.:

Das Eppendorf Analysesystem.

Sonderdruck aus: Das ärztliche Laboratorium 15 (1969) 86

3 Bock, H. H.:

Automatische Klassifikation. Theoretische und praktische Methoden
zur Gruppierung und Strukturierung von Daten (Cluster-Analyse).
Vandenhoek u. Rupprecht, Göttingen (1965)

4 Boehringer Mannheim:

Qualitätskontrolle. Unser Programm - Kontrollseren und Kontroll-
plasmen von Boehringer Mannheim zur Präzisions- und Richtigkeits-
kontrolle.

5 Boroviczeny, K. G. v., R. Merten:

Systematik der Qualitätskontrolle im medizinischen Laboratorium.

6 Büttner, H.:

Qualitätskontrolle im automatisierten klinisch-chemischen Labora-
torium.

Z. klin. Chem. u. klin. Biochemie (1969) 89

- 7 Büttner, H.:
Qualitätskontrolle im klinisch-chemischen Routinelaboratorium.
Z. anal. Chem. 212 (1965) 109
- 8 Büttner, H., E. Hansert, D. Stamm:
Auswertung, Kontrolle und Beurteilung von Meßergebnissen.
In Bergmeyer, H.: Methoden der enzymatischen Analyse. 2. Auflage,
Verlag Chemie, Weinheim Bergstraße (1970) 281
- 9 Cederblad, G., et al.:
Computer-Assisted Evaluation by Multivariate Reference Region
and Cluster Analysis as Applied to Serum Thyroid Hormones and
Thyrotropin.
In Barber, B., Gremy, F., Ueberla, K., Wagner, G. (Eds.): Medical
Informatics Berlin (1979), Springer-Verlag Berlin - Heidelberg -
New York (1979)
- 10 Cramer, H.:
Mathematical methods of statistics.
Princeton University Press (1970)
- 11 Glick, J. H.:
Statistics of Patient Test Values: Application to Indirect Normal
Range and to Quality Control.
Clinical Chemistry 18 (1972) 1504
- 12 Gowenlock, A. H., P. N. C. Broughton:
The Influence of Accuracy and Precision on the Normal Range and
on Acceptable Limits for an Analytical Result.
Z. anal. Chem. 243 (1968) 774

- 13 Grahau, D. F., G. Kenny, R. Wright:
Computer Diagnosis of Acute Abdominal Pain.
In Barber, B., Gremy, F., Ueberla, K., Wagner, G. (Eds.): Medical Informatics Berlin (1979), Springer-Verlag Berlin - Heidelberg - New York (1979)
- 14 Griesser, G., G. Wagner:
Automatisierung des klinischen Laboratoriums. Dokumentation und Statistik von Laboratoriumsergebnissen und medizinisch-technischen Daten.
Bericht über die 12. Jahrestagung der Deutschen Gesellschaft für Medizinische Dokumentation und Statistik in der DGD e.V. vom 09. bis 11. Okt. 1967 in Kiel. Schattauer-Verlag (1968)
- 15 Hilden, H., J. D. F. Habbema, B. Bjerregaard:
The Measurement of Performance in Probabilistic Diagnosis.
Meth. Inform. Med., Vol. 17, No. 4 (1978) 238
- 16 Jacobs, J. F., et al.:
Computer Diagnosis generated from SMA-12/60 chemistry Profiles.
Journal of the Association for the Advancement of Medical Instrumentation, Vol. 6, No. 1, February (1972)
- 17 Kabasawa, K., et al.:
Sequential Diagnostic System Based on Reduction Theory.
In Anderson, J., Forsythe, J. M. (Eds.): MEDINFO 74, Vol. 1, North-Holland, Amsterdam (1975)
- 18 Kissjov, V. T., D. A. Nenov:
A Method for Correlation Analysis of Medical Information.
In Shires, D. B., Wolff, H. (Eds.): MEDINFO 77, North-Holland, Amsterdam (1977)

- 19 Kissjov, V. T., M. E. Aladjemov:
A Method of Determination of Linear Symptom Combinations
Applied to Differential Diagnosis between Cerebral Haemorrhagy
and Infarction.
In Barber, B., Gremy, F., Ueberla, K., Wagner, G. (Eds.): Medical
Informatics Berlin (1979), Springer-Verlag Berlin - Heidelberg -
New York (1979)
- 20 Koller, S.:
Die Problematik der Normalwerte in der Medizin.
Bericht über die 12. Jahrestagung der Deutschen Gesellschaft für
Medizinische Dokumentation und Statistik in der DGD e.V. vom 09.
bis 11. Okt. 1967 in Kiel. Schattauer-Verlag (1968) 141
- 21 Kreyszig, E.:
Statistische Methoden und ihre Anwendung, 4. Auflage.
Vandenhoeck u. Rupprecht, Göttingen (1973)
- 22 Leminor, M., A. Alperovitch, et al.:
Strategy for the Diagnosis of Hepato-Biliary Disease: The Computer-
Aided Decision.
In Medcomp 77, Online Conferences Limited, Uxbridge, England
- 23 McCracken, D. D.:
Fortran.
Karl Hanser Verlag, München (1970)
- 24 Merten, R., K. G. v. Boroviczeny:
Fehlermöglichkeiten bei Laboratoriumswerten.
Diagnostik 5 (1972) 172

- 25 Merten, R., K. G. v. Boroviczeny:
Zufällige und systematische Fehler im Laboratorium.
Diagnostik 5 (1972) 325
- 26 Mood, A., F. Graybill, D. Boes: Introduction to the Theorie of
Statistics.
McGraw-Hill (1974) third edition
- 27 Nakacke, J. P.:
What is Data Analysis and its Current Medical Applications?
In Medcomp 77, Online Conferences Limited, Uxbridge, England
- 28 Niemann, H.:
Methoden der Mustererkennung.
Akademische Verlagsgesellschaft, Frankfurt am Main (1974)
- 29 Nomura, Y., S. Shimmura:
Computer Assisted Prognosis of Acute Myocardial Infarction.
In Shires, D. B., Wolff, H. (Eds.) MEDINFO 77, North-Holland,
Amsterdam (1977)
- 30 Ortiz-Berrocal, J., et al.:
Discriminant analysis and Bayesian method in the diagnostic
decision. Application in thyreoid pathology.
In Medcomp 77, Online Conferences Limited, Uxbridge, England
- 31 Owen, J., D. Campbell:
A comparison of plasma electrolyte and urea values in healthy
persons and in hospital patients.
Clin. Chim. Acta 22 (1968) 611

- 32 Patrick, E. A., et al.:
Pattern Recognition Applied to Early Diagnosis of Heart Attack.
In Shires, D. B., Wolff, H. (Eds.): MEDINFO 77, North-Holland,
Amsterdam (1977)
- 33 Paul, L. F.:
Sequencing Laboratory Tests for Improved Sequential Diagnosis.
An Anderson, J., Forsythe, J. M. (Eds.): MEDINFO 74, Vol. 1,
North-Holland, Amsterdam (1975)
- 34 Reichertz, P. L.:
Computer Aided Medical Practice Oriented towards Diagnosis.
In Shires, D. B., Wolff, H. (Eds.): MEDINFO 77, North-Holland,
Amsterdam (1977)
- 35 Richtlinien der Bundesärztekammer zur Durchführung der statisti-
schen Qualitätskontrolle und von Ringversuchen im Bereich der
Heilkunde.
- 36 Roberts, L. B.:
Clin. Chim. Acta 16 (1967) 69
- 37 Suchet, A.:
Statistische Qualitätskontrolle einer manuellen Mikroliter-Serien-
methode.
Das medizinische Laboratorium, Jahrgang 23, Heft 2, Seite 32
- 38 Szasz, G.:
Die Qualitätskontrolle von Enzymaktivitätsbestimmungen im Serum.
Z. klin. Chem. u. klin. Biochem. 8 (1970) 212

39 Stamm, D.:

Tagesschwankungen der Konzentration diagnostisch wichtiger
Blutbestandteile.

Habilitationsschrift, Gießen (1966) 65

40 Thurnmayr, R., et al.:

Computer Aid for Diagnosing Pancreatic, Hepatic and Gastric
Diseases by Function Test.

In Anderson, J., Forsythe, J. M. (Eds.): MEDINFO 74, Vol. 1,
North-Holland, Amsterdam (1975)

41 Tonks, D. B.:

Clin. Chem. 2 (1956) 296

42 Waerden, B. L. v. d.:

Mathematische Statistik, 3. Auflage.

Springer-Verlag Berlin (1971)

43 Walter, G. F., Ch. J. Kowalski:

Computer Aided Diagnosis of Craniofacial Abnormalities.

In Anderson, J., Forsythe, J. M. (Eds.): MEDINFO 74, Vol. 1,
North-Holland, Amsterdam (1975)

44 Wissenschaftliche Tabellen von Geigy

Herrn Prof. Dr. L. Horbach danke ich für die Möglichkeit der Erstellung dieser Dissertation.

Mein besonderer Dank gilt Frau Dr. rer. nat. Dr. med. I. Guggenmoos-Holzmann für wertvolle Anregungen.

LEBENS LAUF

- Am 27. 05. 1950 wurde ich in Schweinfurt/Main als Sohn des
Diplom-Forstwirtes Herbert Münch und seiner
Frau Franziska Münch geboren.
- Von 1956 bis 1960 besuchte ich die Volksschule in Würzburg.
- Von 1960 bis 1969 wurde ich an den Humanistischen Gymnasien
von Würzburg, Aschaffenburg und Ansbach unter-
richtet. An letzterem absolvierte ich 1969 das
Abitur als Klassenbester. Nach einer Prüfung
am Melanchton-Gymnasium in Nürnberg erhielt
ich das Bayerische Begabten-Stipendium.
- Von 1969 bis 1970 leistete ich meinen Wehrdienst.
- 1970 begann ich an der Universität Erlangen mit dem
Studium der Informatik,
- 1971 mit dem Studium der Humanmedizin.
- 1978 absolvierte ich die Ärztliche Prüfung und
- 1979 die Informatik-Hauptdiplomprüfung.
- Vom 01. 12. 1978
bis 31. 12. 1979 war ich wissenschaftlicher Mitarbeiter am
Institut für Medizinische Informatik der Medi-
zinischen Hochschule Hannover.
- Seit 01. 01. 1980 bin ich wissenschaftlicher Mitarbeiter der
Universitäts-Hals-Nasen-Ohren-Klinik in Erlangen.

Erlangen, im November 1981

Erwin Münch

Copy-Center 2000

852 ERLANGEN · Obere Karlstr. 30
Tel. 09131-26197